# Программирование сокетов в Linux

Socket API был впервые реализован в операционной системе Berkley UNIX. Сейчас этот программный интерфейс доступен практически в любой модификации Unix, в том числе в Linux. Хотя все реализации чем-то отличаются друг от друга, основной набор функций в них совпадает. Изначально сокеты использовались в программах на C/C++, но в настоящее время средства для работы с ними предоставляют многие языки (C#, Java и др.).

Сокеты предоставляют весьма мощный и гибкий механизм межпроцессного взаимодействия (IPC). Они могут использоваться для организации взаимодействия программ на одном компьютере, по локальной сети или через Internet, что позволяет создавать распределённые приложения различной сложности. Кроме того, с их помощью можно организовать взаимодействие с программами, работающими под управлением других операционных систем. Например, под Windows существует интерфейс Window Sockets, спроектированный на основе socket API.

Сокеты поддерживают многие стандартные сетевые протоколы (конкретный их список зависит от реализации) и предоставляют унифицированный интерфейс для работы с ними. Наиболее часто сокеты используются для работы в IP-сетях. В этом случае их можно использовать для взаимодействия приложений не только по специально разработанным, но и по стандартным протоколам - HTTP, FTP, Telnet и т. д. Например, c использованием сокетов можно написать собственный Web-броузер или Web-сервер, способный обслуживать одновременно множество клиентов.

Сокеты - весьма мощное и удобное средство для сетевого программирования.

# Основы socket API

Сокет (socket) - это конечная точка сетевых коммуникаций. Он является чем-то вроде "портала", через которое можно отправлять байты во внешний мир. Приложение просто пишет данные в сокет; их дальнейшая буферизация, отправка и транспортировка осуществляется используемым стеком протоколов и сетевой аппаратурой. Чтение данных из сокета происходит аналогичным образом.

В программе сокет идентифицируется **дескриптором** - это просто переменная типа int. Программа получает дескриптор от операционной системы при создании сокета, а затем передаёт его сервисам socket API для указания сокета, над которым необходимо выполнить то или иное действие.

С каждым сокет связываются **три атрибута: домен, тип и протокол**. Эти атрибуты задаются при создании сокета и остаются неизменными на протяжении всего времени его существования. Для создания сокета используется функция socket, имеющая следующий прототип.

#include <sys/types.h>

#include <sys/socket.h>

int socket(int domain, int type, int protocol);

Домен определяет пространство адресов, в котором располагается сокет, и множество протоколов, которые используются для передачи данных. Чаще других используются домены Unix и Internet, задаваемые константами AF\_UNIX и AF\_INET соответственно (префикс AF означает "address family" - "семейство адресов"). При задании AF\_UNIX для передачи данных используется файловая система ввода/вывода Unix. В этом случае сокеты используются для межпроцессного взаимодействия на одном компьютере и не годятся для работы по сети. Константа AF\_INET соответствует Internet-домену. Сокеты, размещённые в этом домене, могут использоваться для работы в любой IP-сети. Существуют и другие домены (AF\_IPX для протоколов Novell, AF\_INET6 для модификации протокола IP - IPv6 и т. д.).

**Тип сокета** определяет способ передачи данных по сети. Чаще других применяются:

**SOCK\_STREAM**. Передача потока данных с предварительной установкой соединения. Обеспечивается надёжный канал передачи данных, при котором фрагменты отправленного блока не теряются, не переупорядочиваются и не дублируются.

**SOCK\_DGRAM**. Передача данных в виде отдельных сообщений (датаграмм). Предварительная установка соединения не требуется. Обмен данными происходит быстрее, но является ненадёжным: сообщения могут теряться в пути, дублироваться и переупорядочиваться. Допускается передача сообщения нескольким получателям (multicasting) и широковещательная передача (broadcasting).

**SOCK\_RAW**. Этот тип присваивается низкоуровневым сокетам. Их отличие от обычных сокетов состоит в том, что с их помощью программа может взять на себя формирование некоторых заголовков, добавляемых к сообщению.

Обратите внимание, что не все домены допускают задание произвольного типа сокета. Например, совместно с доменом Unix используется только тип SOCK\_STREAM. С другой стороны, для Internet-домена можно задавать любой из перечисленных типов. В этом случае для реализации SOCK\_STREAM используется протокол TCP, для реализации SOCK\_DGRAM - протокол UDP, а тип SOCK\_RAW используется для низкоуровневой работы с протоколами IP, ICMP и т. д.

Наконец, последний атрибут определяет **протокол**, используемый для передачи данных. Часто протокол однозначно определяется по домену и типу сокета. В этом случае в качестве третьего параметра функции socket можно передать 0, что соответствует протоколу по умолчанию. Тем не менее, иногда (например, при работе с низкоуровневыми сокетами) требуется задать протокол явно. Числовые идентификаторы протоколов зависят от выбранного домена.

# Адреса

Прежде чем передавать данные через сокет, его необходимо связать с адресом в выбранном домене (эту процедуру называют именованием сокета). Иногда связывание осуществляется неявно (внутри функций connect и accept), но выполнять его необходимо во всех случаях. Вид адреса зависит от выбранного вами домена. В Unix-домене это текстовая строка - имя файла, через который происходит обмен данными. В Internet-домене адрес задаётся комбинацией IP-адреса и 16-битного номера порта. IP-адрес определяет хост в сети, а порт - конкретный сокет на этом хосте. Протоколы TCP и UDP используют различные пространства портов.

Для явного связывания сокета с некоторым адресом используется функция bind. Её прототип имеет вид:

#include <sys/types.h>

#include <sys/socket.h>

int bind(int sockfd, struct sockaddr \*addr, int addrlen);

В качестве первого параметра передаётся дескриптор сокета, который мы хотим привязать к заданному адресу. Второй параметр, addr, содержит указатель на структуру с адресом, а третий - длину этой структуры.

struct sockaddr {

unsigned short sa\_family; // Семейство адресов, AF\_xxx

char sa\_data[14]; // 14 байтов для хранения адреса

};

Поле sa\_family содержит идентификатор домена, тот же, что и первый параметр функции socket. В зависимости от значения этого поля по-разному интерпретируется содержимое массива sa\_data. Разумеется, работать с этим массивом напрямую не очень удобно, поэтому можно использовать вместо sockaddr одну из альтернативных структур вида sockaddr\_XX (XX - суффикс, обозначающий домен: "un" - Unix, "in" - Internet и т. д.). При передаче в функцию bind указатель на эту структуру приводится к указателю на sockaddr. Рассмотрим для примера структуру sockaddr\_in.

struct sockaddr\_in {

short int sin\_family; // Семейство адресов

unsigned short int sin\_port; // Номер порта

struct in\_addr sin\_addr; // IP-адрес

unsigned char sin\_zero[8]; // "Дополнение" до размера структуры sockaddr

};

Здесь поле sin\_family соответствует полю sa\_family в sockaddr, в sin\_port записывается номер порта, а в sin\_addr - IP-адрес хоста. Поле sin\_addr само является структурой, которая имеет вид:

struct in\_addr {

unsigned long s\_addr;

};

Зачем понадобилось заключать всего одно поле в структуру? Дело в том, что раньше in\_addr представляла собой объединение (union), содержащее гораздо большее число полей. Сейчас, когда в ней осталось всего одно поле, она продолжает использоваться для обратной совместимости.

**ПРИМЕЧАНИЕ**

Существует два порядка хранения байтов в слове и двойном слове. Один из них называется порядком хоста (host byte order), другой - сетевым порядком (network byte order) хранения байтов. При указании IP-адреса и номера порта необходимо преобразовать число из порядка хоста в сетевой. Для этого используются функции htons (Host TO Network Short) и htonl (Host TO Network Long). Обратное преобразование выполняют функции ntohs и ntohl.

На некоторых машинах (к PC это не относится) порядок хоста и сетевой порядок хранения байтов совпадают. Тем не менее, функции преобразования лучше применять и там, поскольку это улучшит переносимость программы. Это никак не скажется на производительности, так как препроцессор сам уберёт все "лишние" вызовы этих функций, оставив их только там, где преобразование действительно необходимо.

# Установка соединения (сервер)

Установка соединения на стороне сервера состоит из четырёх этапов, ни один из которых не может быть опущен. Сначала сокет создаётся и привязывается к локальному адресу. Если компьютер имеет несколько сетевых интерфейсов с различными IP-адресами, вы можете принимать соединения только с одного из них, передав его адрес функции bind. Если же вы готовы соединяться с клиентами через любой интерфейс, задайте в качестве адреса константу INADDR\_ANY. Что касается номера порта, вы можете задать конкретный номер или 0 (в этом случае система сама выберет произвольный неиспользуемый в данный момент номер порта).

На следующем шаге создаётся очередь запросов на соединение. При этом сокет переводится в режим ожидания запросов со стороны клиентов. Всё это выполняет функция listen.

int listen(int sockfd, int backlog);

Первый параметр - дескриптор сокета, а второй задаёт размер очереди запросов. Каждый раз, когда очередной клиент пытается соединиться с сервером, его запрос ставится в очередь, так как сервер может быть занят обработкой других запросов. Если очередь заполнена, все последующие запросы будут игнорироваться. Когда сервер готов обслужить очередной запрос, он использует функцию accept.

#include <sys/socket.h>

int accept(int sockfd, void \*addr, int \*addrlen);

Функция accept создаёт для общения с клиентом новый сокет и возвращает его дескриптор. Параметр sockfd задаёт слушающий сокет. После вызова он остаётся в слушающем состоянии и может принимать другие соединения. В структуру, на которую ссылается addr, записывается адрес сокета клиента, который установил соединение с сервером. В переменную, адресуемую указателем addrlen, изначально записывается размер структуры; функция accept записывает туда длину, которая реально была использована. Если вас не интересует адрес клиента, вы можете просто передать NULL в качестве второго и третьего параметров.

Следует обратить внимание, что полученный от accept новый сокет связан с тем же самым адресом, что и слушающий сокет. Сначала это может показаться странным. Но дело в том, что адрес TCP-сокета не обязан быть уникальным в Internet-домене. Уникальными должны быть только соединения, для идентификации которых используются два адреса сокетов, между которыми происходит обмен данными.

# Установка соединения (клиент)

На стороне клиента для установления соединения используется функция connect, которая имеет следующий прототип.

#include <sys/types.h>

#include <sys/socket.h>

int connect(int sockfd, struct sockaddr \*serv\_addr, int addrlen);

Здесь sockfd - сокет, который будет использоваться для обмена данными с сервером, serv\_addr содержит указатель на структуру с адресом сервера, а addrlen - длину этой структуры. Обычно сокет не требуется предварительно привязывать к локальному адресу, так как функция connect сделает это за вас, подобрав подходящий свободный порт. Вы можете принудительно назначить клиентскому сокету некоторый номер порта, используя bind перед вызовом connect. Делать это следует в случае, когда сервер соединяется с только с клиентами, использующими определённый порт (примерами таких серверов являются rlogind и rshd). В остальных случаях проще и надёжнее предоставить системе выбрать порт за вас.

# Обмен данными

После того как соединение установлено, можно начинать обмен данными. Для этого используются функции send и recv. В Unix для работы с сокетами можно использовать также файловые функции read и write, но они обладают меньшими возможностями, и, кроме того, не будут работать на других платформах (например, под Windows).

Функция send используется для отправки данных и имеет следующий прототип:

int send(int sockfd, const void \*msg, int len, int flags);

Здесь sockfd - это, как всегда, дескриптор сокета, через который мы отправляем данные, msg - указатель на буфер с данными, len - длина буфера в байтах, а flags - набор битовых флагов, управляющих работой функции (если флаги не используются, передайте функции 0). Вот некоторые из них (полный список можно найти в документации):

**MSG\_OOB**. Предписывает отправить данные как срочные (out of band data, OOB). Концепция срочных данных позволяет иметь два параллельных канала данных в одном соединении. Иногда это бывает удобно. Например, Telnet использует срочные данные для передачи команд типа Ctrl+C. В настоящее время использовать их не рекомендуется из-за проблем с совместимостью (существует два разных стандарта их использования, описанные в RFC793 и RFC1122). Безопаснее просто создать для срочных данных отдельное соединение.

**MSG\_DONTROUTE**. Запрещает маршрутизацию пакетов. Нижележащие транспортные слои могут проигнорировать этот флаг.

Функция send возвращает число байтов, которое на самом деле было отправлено (или -1 в случае ошибки). Это число может быть меньше указанного размера буфера. Если вы хотите отправить весь буфер целиком, вам придётся написать свою функцию и вызывать в ней send, пока все данные не будут отправлены. Она может выглядеть примерно так.

int sendall(int s, char \*buf, int len, int flags)

{

int total = 0;

int n;

while(total < len)

{

n = send(s, buf+total, len-total, flags);

if(n == -1) { break; }

total += n;

}

return (n==-1 ? -1 : total);

}

Использование sendall ничем не отличается от использования send, но она отправляет весь буфер с данными целиком.

Для чтения данных из сокета используется функция recv.

int recv(int sockfd, void \*buf, int len, int flags);

В целом её использование аналогично send. Она точно так же принимает дескриптор сокета, указатель на буфер и набор флагов. Флаг MSG\_OOB используется для приёма срочных данных, а MSG\_PEEK позволяет "подсмотреть" данные, полученные от удалённого хоста, не удаляя их из системного буфера (это означает, что при следующем обращении к recv вы получите те же самые данные). По аналогии с send функция recv возвращает количество прочитанных байтов, которое может быть меньше размера буфера. Существует один особый случай, при котором recv возвращает 0. Это означает, что соединение было разорвано.

# Закрытие сокета

Закончив обмен данными сокет необходимо закрыть с помощью функции close. Это приведёт к разрыву соединения.

#include <unistd.h>

int close(int fd);

Вы также можете запретить передачу данных в каком-то одном направлении, используя функцию shutdown.

int shutdown(int sockfd, int how);

Параметр how может принимать одно из следующих значений:

0 - запретить чтение из сокета

1 - запретить запись в сокет

2 - запретить и то и другое

Хотя после вызова shutdown с параметром how, равным 2, вы больше не сможете использовать сокет для обмена данными, вам всё равно потребуется вызвать close, чтобы освободить связанные с ним системные ресурсы.

# Обработка ошибок

Если в процессе обмена данными что-то пошло не так, все рассмотренные нами функции возвращают -1, записывая в глобальную переменную errno код ошибки. Соответственно, вы можете проанализировать значение этой переменной и предпринять действия по восстановлению нормальной работы программы, не прерывая её выполнения. А можете просто выдать диагностическое сообщение (для этого удобно использовать функцию perror), а затем завершить программу с помощью exit.

# Отладка программ

Как можно отлаживать сетевую программу, если под рукой нет сети. Оказывается, можно обойтись и без неё. Достаточно запустить клиента и сервера на одной машине, а затем использовать для соединения адрес интерфейса внутренней петли (loopback interface). В программе ему соответствует константа INADDR\_LOOPBACK (не забудьте применять к ней функцию htonl!). Пакеты, направляемые по этому адресу, в сеть не попадают. Вместо этого они передаются стеку протоколов TCP/IP как только что принятые. Таким образом моделируется наличие виртуальной сети, в которой вы можете отлаживать ваши сетевые приложения.

# Практический пример эхо-клиент и эхо-сервер

Теперь, когда мы изучили основные функции для работы с сокетами, самое время посмотреть, как они используются на практике. Эхо-клиент посылает сообщение "Hello there!" и выводит на экран ответ сервера. Его код приведён в листинге 1. Эхо-сервер читает всё, что передаёт ему клиент, а затем просто отправляет полученные данные обратно. Его код содержится в листинге 2.

Листинг 1. Эхо-клиент.

#include <sys/types.h>

#include <sys/socket.h>

#include <netinet/in.h>

char message[] = "Hello there!\n";

char buf[sizeof(message)];

int main()

{

int sock;

struct sockaddr\_in addr;

sock = socket(AF\_INET, SOCK\_STREAM, 0);

if(sock < 0)

{

perror("socket");

exit(1);

}

addr.sin\_family = AF\_INET;

addr.sin\_port = htons(3425); // или любой другой порт...

addr.sin\_addr.s\_addr = htonl(INADDR\_LOOPBACK);

if(connect(sock, (struct sockaddr \*)&addr, sizeof(addr)) < 0)

{

perror("connect");

exit(2);

}

send(sock, message, sizeof(message), 0);

recv(sock, buf, sizeof(message), 0);

printf(buf);

close(sock);

return 0;

}

Листинг 2. Эхо-сервер.

#include <sys/types.h>

#include <sys/socket.h>

#include <netinet/in.h>

int main()

{

int sock, listener;

struct sockaddr\_in addr;

char buf[1024];

int bytes\_read;

listener = socket(AF\_INET, SOCK\_STREAM, 0);

if(listener < 0)

{

perror("socket");

exit(1);

}

addr.sin\_family = AF\_INET;

addr.sin\_port = htons(3425);

addr.sin\_addr.s\_addr = htonl(INADDR\_ANY);

if(bind(listener, (struct sockaddr \*)&addr, sizeof(addr)) < 0)

{

perror("bind");

exit(2);

}

listen(listener, 1);

while(1)

{

sock = accept(listener, NULL, NULL);

if(sock < 0)

{

perror("accept");

exit(3);

}

while(1)

{

bytes\_read = recv(sock, buf, 1024, 0);

if(bytes\_read <= 0) break;

send(sock, buf, bytes\_read, 0);

}

close(sock);

}

return 0;

}

# Обмен датаграммами

Как уже говорилось, датаграммы используются в программах довольно редко. В большинстве случаев надёжность передачи критична для приложения, и вместо изобретения собственного надёжного протокола поверх UDP программисты предпочитают использовать TCP. Тем не менее, иногда датаграммы оказываются полезны. Например, их удобно использовать при транслировании звука или видео по сети в реальном времени, особенно при широковещательном транслировании.

Поскольку для обмена датаграммами не нужно устанавливать соединение, использовать их гораздо проще. Создав сокет с помощью socket и bind, вы можете тут же использовать его для отправки или получения данных. Для этого вам понадобятся функции sendto и recvfrom.

int sendto(int sockfd, const void \*msg, int len, unsigned int flags,

const struct sockaddr \*to, int tolen);

int recvfrom(int sockfd, void \*buf, int len, unsigned int flags,

struct sockaddr \*from, int \*fromlen);

Функция sendto очень похожа на send. Два дополнительных параметра to и tolen используются для указания адреса получателя. Для задания адреса используется структура sockaddr, как и в случае с функцией connect. Функция recvfrom работает аналогично recv. Получив очередное сообщение, она записывает его адрес в структуру, на которую ссылается from, а записанное количество байт - в переменную, адресуемую указателем fromlen. Как мы знаем, аналогичным образом работает функция accept.

Некоторую путаницу вносят присоединённые датаграммные сокеты (connected datagram sockets). Дело в том, что для сокета с типом SOCK\_DGRAM тоже можно вызвать функцию connect, а затем использовать send и recv для обмена данными. Нужно понимать, что никакого соединения при этом не устанавливается. Операционная система просто запоминает адрес, который вы передали функции connect, а затем использует его при отправке данных. Обратите внимание, что присоединённый сокет может получать данные только от сокета, с которым он соединён.

Для иллюстрации процесса обмена датаграммами расссмотрим две небольшие программы - sender (листинг 3) и receiver (листинг 4). Первая отправляет сообщения "Hello there!" и "Bye bye!", а вторая получает их и печатает на экране. Программа sender демонстрирует применение как обычного, так и присоединённого сокета, а receiver использует обычный.

Листинг 3. Программа sender.

#include <sys/types.h>

#include <sys/socket.h>

#include <netinet/in.h>

char msg1[] = "Hello there!\n";

char msg2[] = "Bye bye!\n";

int main()

{

int sock;

struct sockaddr\_in addr;

sock = socket(AF\_INET, SOCK\_DGRAM, 0);

if(sock < 0)

{

perror("socket");

exit(1);

}

addr.sin\_family = AF\_INET;

addr.sin\_port = htons(3425);

addr.sin\_addr.s\_addr = htonl(INADDR\_LOOPBACK);

sendto(sock, msg1, sizeof(msg1), 0,

(struct sockaddr \*)&addr, sizeof(addr));

connect(sock, (struct sockaddr \*)&addr, sizeof(addr));

send(sock, msg2, sizeof(msg2), 0);

close(sock);

return 0;

}

Листинг 4. Программа receiver.

#include <sys/types.h>

#include <sys/socket.h>

#include <netinet/in.h>

#include <stdio.h>

int main()

{

int sock;

struct sockaddr\_in addr;

char buf[1024];

int bytes\_read;

sock = socket(AF\_INET, SOCK\_DGRAM, 0);

if(sock < 0)

{

perror("socket");

exit(1);

}

addr.sin\_family = AF\_INET;

addr.sin\_port = htons(3425);

addr.sin\_addr.s\_addr = htonl(INADDR\_ANY);

if(bind(sock, (struct sockaddr \*)&addr, sizeof(addr)) < 0)

{

perror("bind");

exit(2);

}

while(1)

{

bytes\_read = recvfrom(sock, buf, 1024, 0, NULL, NULL);

buf[bytes\_read] = '\0';

printf(buf);

}

return 0;

}

# Использование низкоуровневых сокетов

Низкоуровневые сокеты открывают перед вами новые горизонты. Они предоставляют программисту полный контроль над содержимым пакетов, которые отправляются в путешествие по сети. С другой стороны, они сложнее в использовании и обладают плохой переносимостью. Вот почему использовать их следует только в случае необходимости. Например, без них не обойтись при разработке системных утилит типа ping и traceroute.

Работая с обычными сокетами, вы передаёте системе "чистые" данные, а она сама заботится о добавлении к ним необходимых заголовков (а иногда ещё и концевиков). Например, когда вы посылаете сообщение через UDP-сокет, к нему добавляется сначала UDP-заголовок, потом IP-заголовок, а в самом конце - заголовок аппаратного протокола, который используется в вашей локальной сети (например, Ethernet). В результате получается кадр, показанный на рисунке 1.



Рисунок 1

Низкоуровневые сокеты позволяют вам включать в буфер с данными заголовки некоторых протоколов. Например, вы можете включить в ваше сообщение TCP- или UDP-заголовок, предоставив системе сформировать для вас IP-заголовок, а можете вообще сформировать все заголовки самостоятельно. Разумеется, при этом вам придётся изучить работу соответствующих протоколов и строго соблюсти формат их заголовков, иначе программа работать не будет.

При работе с низкоуровневыми сокетами вам придётся указывать в третьем параметре функции socket тот протокол, к заголовкам которого вы хотите получить доступ. Константы для основных протоколов Internet объявлены в файле netinet/in.h. Они имеют вид IPPROTO\_XXX, где XXX-название протокола: IPPROTO\_TCP, IPPROTO\_UDP, IPPROTO\_RAW (в последнем случае вы получите возможность поработать с "сырым" IP и формировать IP-заголовки вручную).

Все числовые данные в заголовках должны записываться в сетевом формате. Поэтому не забывайте использовать функции htons и htonl.

Рассмотрим пример с использованием низкоуровневых UDP-сокетов. Для этого необходимо вручную формировать UDP-заголовок отправляемого сообщения (рисунок 2).



Рисунок 2

Код примера приведён в листинге 5. Следует обратить внимание на несколько моментов. Во-первых, в структуре sockaddr\_in не задается номер порта. Поскольку этот номер содержится в UDP-заголовке, от поля sin\_port уже ничего не зависит. Во-вторых, в качестве контрольной суммы взят ноль, чтобы не заниматься её вычислением. Протокол UDP является ненадёжным по своей природе, поэтому он допускает подобную вольность. Но другие протоколы (например, IP) могут и не допускать. Наконец, обратите внимание, что все данные UDP-заголовка форматируются с использованием htons.

Листинг 5. Программа sender с использованием низкоуровневых сокетов.

#include <sys/types.h>

#include <sys/socket.h>

#include <netinet/in.h>

struct UdpHeader

{

u\_short src\_port;

u\_short targ\_port;

u\_short length;

u\_short checksum;

};

char message[] = "Hello there!\n";

char msgbuf[1024];

int main()

{

int sock;

struct sockaddr\_in addr;

struct UdpHeader header;

sock = socket(AF\_INET, SOCK\_RAW, IPPROTO\_UDP);

if(sock < 0)

{

perror("socket");

exit(1);

}

addr.sin\_family = AF\_INET;

addr.sin\_addr.s\_addr = htonl(INADDR\_LOOPBACK);

header.targ\_port = htons(3425);

header.length = htons(sizeof(header)+sizeof(message));

header.checksum = 0;

memcpy((void \*)msgbuf, (void \*)&header, sizeof(header));

memcpy((void \*)(msgbuf+sizeof(header)), (void \*)message, sizeof(message));

sendto(sock, msgbuf, sizeof(header)+sizeof(message), 0,

(struct sockaddr \*)&addr, sizeof(addr));

close(sock);

return 0;

}

# Функции для работы с адресами и DNS

Поскольку для идентификации хостов в Internet широко используются доменные имена, нам необходимо изучить механизм преобразования их в IP-адреса.

IP-адреса принято записывать в виде четырёх чисел, разделённых точками. Для преобразования адреса, записанного в таком формате, в число и наоборот используется семейство функций inet\_addr, inet\_aton и inet\_ntoa.

#include <sys/socket.h>

#include <netinet/in.h>

#include <arpa/inet.h>

int inet\_aton(const char \*cp, struct in\_addr \*in\_p);

unsigned long int inet\_addr(const char \*cp);

char \*inet\_ntoa(struct in\_addr in);

Функция inet\_addr часто используется в программах. Она принимает строку и возвращает адрес (уже с сетевым порядком следования байтов). Проблема с этой функцией состоит в том, что значение -1, возвращаемое ею в случае ошибки, является в то же время корректным адресом 255.255.255.255 (широковещательный адрес). Вот почему рекомендуется использовать функцию inet\_aton (Ascii TO Network). Для обратного преобразования используется функция inet\_ntoa (Network TO Ascii). Обе эти функции работают с адресами в сетевом формате. Обратите внимание, что в случае ошибки они возвращают 0, а не -1.

Для преобразования доменного имени в IP-адрес используется функция gethostbyname.

#include <netdb.h>

struct hostent \*gethostbyname(const char \*name);

Эта функция получает имя хоста и возвращает указатель на структуру с его описанием. Рассмотрим эту структуру более подробно.

struct hostent {

char \*h\_name;

char \*\*h\_aliases;

int h\_addrtype;

int h\_length;

char \*\*h\_addr\_list;

};

#define h\_addr h\_addr\_list[0]

* h\_name. Имя хоста.
* h\_aliases. Массив строк, содержащих псевдонимы хоста. Завершается значением NULL.
* h\_addrtype. Тип адреса. Для Internet-домена - AF\_INET.
* h\_length. Длина адреса в байтах.
* h\_addr\_list. Массив, содержащий адреса всех сетевых интерфейсов хоста. Завершается нулём. Обратите внимание, что байты каждого адреса хранятся с сетевым порядке, поэтому htonl вызывать не нужно.

Как видим, gethostbyname возвращает достаточно полную информацию. Если нас интересует адрес хоста, мы можем выбрать его из массива h\_addr\_list. Часто берут самый первый адрес (как мы видели выше, для ссылки на него определён специальный макрос h\_addr). Для определения имени хоста по адресу используется функция gethostbyaddr. Вместо строки она получает адрес (в виде sockaddr) и возвращает указатель на ту же самую структуру hostent. Используя эти две функции, нужно помнить, что они сообщают об ошибке не так, как остальные: вместо указателя возвращается NULL, а расширенный код ошибки записывается в глобальную переменную h\_errno (а не errno). Соответственно, для вывода диагностического сообщения следует использовать herror вместо perror.

Следует иметь в виду, что функции gethostbyname и gethostbyaddr возвращают указатель на статическую область памяти. Это означает, что каждое новое обращение к одной из этих функций приведёт к перезаписи данных, полученных при преыдущем обращении.

В заключение рассмотрим ещё одно семейство полезных функций - gethostname, getsockname и getpeername.

#include <unistd.h>

int gethostname(char \*hostname, size\_t size);

Функция gethostname используется для получения имени локального хоста. Далее его можно преобразовать в адрес при помощи gethostbyname. Это даёт нам способ в любой момент программно получить адрес машины, на которой выполняется наша программа, что может быть полезным во многих случаях.

#include <sys/socket.h>

int getpeername(int sockfd, struct sockaddr \*addr, int \*addrlen);

Функция getpeername позволяет в любой момент узнать адрес сокета на "другом конце" соединения. Она получает дескриптор сокета, соединённого с удалённым хостом, и записывает адрес этого хоста в структуру, на которую указывает addr. Фактическое количество записанных байт помещается по адресу addrlen (не забудьте записать туда размер структуры addr до вызова getpeername). Полученный адрес при необходимости можно преобразовать в строку, используя inet\_ntoa или gethostbyaddr. Функция getsockname по назначению обратна getpeername и позволяет определить адрес сокета на "нашем конце" соединения.

# Параллельное обслуживание клиентов

Следующий важный вопрос, который нам предстоит обсудить, - это параллельное обслуживание клиентов. Эта проблема становится актуальной, когда сервер должен обслуживать большое количество запросов. Конечно, на машине с одним процессором настоящей параллельности достичь не удастся. Но даже на одной машине можно добиться существенного выигрыша в производительности. Допустим, сервер отправил какие-то данные клиенту и ждёт подтверждения. Пока оно путешествует по сети, сервер вполне мог бы заняться другими клиентами. Для реализации такого алгоритма обслуживания существует множество способов, но чаще всего применяются два из них.

**Способ 1**

Этот способ подразумевает создание дочернего процесса для обслуживания каждого нового клиента. При этом родительский процесс занимается только прослушиванием порта и приёмом соединений. Чтобы добиться такого поведения, сразу после accept сервер вызывает функцию fork для создания дочернего процесса. Далее анализируется значение, которое вернула эта функция. В родительском процессе оно содержит идентификатор дочернего, а в дочернем процессе равно нулю. Используя этот признак, мы переходим к очередному вызову accept в родительском процессе, а дочерний процесс обслуживает клиента и завершается (\_exit).

С использованием этой методики вышеприведенный эхо-сервер перепишется, как показано в листинге 6.

Листинг 6. Эхо-сервер (версия 2, fork)

#include <sys/types.h>

#include <sys/socket.h>

#include <netinet/in.h>

int main()

{

int sock, listener;

struct sockaddr\_in addr;

char buf[1024];

int bytes\_read;

listener = socket(AF\_INET, SOCK\_STREAM, 0);

if(listener < 0)

{

perror("socket");

exit(1);

}

addr.sin\_family = AF\_INET;

addr.sin\_port = htons(3425);

addr.sin\_addr.s\_addr = INADDR\_ANY;

if(bind(listener, (struct sockaddr \*)&addr, sizeof(addr)) < 0)

{

perror("bind");

exit(2);

}

listen(listener, 1);

while(1)

{

sock = accept(listener, NULL, NULL);

if(sock < 0)

{

perror("accept");

exit(3);

}

switch(fork())

{

case -1:

perror("fork");

break;

case 0:

close(listener);

while(1)

{

bytes\_read = recv(sock, buf, 1024, 0);

if(bytes\_read <= 0) break;

send(sock, buf, bytes\_read, 0);

}

close(sock);

\_exit(0);

default:

close(sock);

}

}

close(listener);

return 0;

}

Очевидное преимущество такого подхода состоит в том, что он позволяет писать весьма компактные, понятные программы, в которых код установки соединения отделён от кода обслуживания клиента. К сожалению, у него есть и недостатки. Во-первых, если клиентов очень много, создание нового процесса для обслуживания каждого из них может оказаться слишком дорогостоящей операцией. Во-вторых, такой способ неявно подразумевает, что все клиенты обслуживаются независимо друг от друга. Однако это может быть не так. Если, к примеру, вы пишете чат-сервер, то ваша основная задача - поддерживать взаимодействие всех клиентов, присоединившихся к нему. В этих условиях границы между процессами станут для вас серьёзной помехой. В подобном случае вам следует серьёзно рассмотреть другой способ обслуживания клиентов.

**Способ 2**

Второй способ основан на использовании неблокирующих сокетов (nonblocking sockets) и функции select. Сокеты, которые мы до сих пор использовали, являлись блокирующими (blocking). Это название означает, что на время выполнения операции с таким сокетом ваша программа блокируется. Например, если вы вызвали recv, а данных на вашем конце соединения нет, то в ожидании их прихода ваша программа "засыпает". Аналогичная ситуация наблюдается, когда вы вызываете accept, а очередь запросов на соединение пуста. Это поведение можно изменить, используя функцию fcntl.

#include <unistd.h>

#include <fcntl.h>

.

.

sockfd = socket(AF\_INET, SOCK\_STREAM, 0);

fcntl(sockfd, F\_SETFL, O\_NONBLOCK);

.

.

Эта несложная операция превращает сокет в неблокирующий. Вызов любой функции с таким сокетом будет возвращать управление немедленно. Причём если затребованная операция не была выполнена до конца, функция вернёт -1 и запишет в errno значение EWOULDBLOCK. Чтобы дождаться завершения операции, мы можем опрашивать все наши сокеты в цикле, пока какая-то функция не вернёт значение, отличное от EWOULDBLOCK. Как только это произойдёт, мы можем запустить на выполнение следующую операцию с этим сокетом и вернуться к нашему опрашивающему циклу. Такая тактика (называемая в англоязычной литературе polling) работоспособна, но очень неэффективна, поскольку процессорное время тратится впустую на многократные (и безрезультатные) опросы.

Чтобы исправить ситуацию, используют функцию select. Эта функция позволяет отслеживать состояние нескольких файловых дескрипторов (а в Unix к ним относятся и сокеты) одновременно.

#include <sys/time.h>

#include <sys/types.h>

#include <unistd.h>

int select(int n, fd\_set \*readfds, fd\_set \*writefds,

fd\_set \*exceptfds, struct timeval \*timeout);

FD\_CLR(int fd, fd\_set \*set);

FD\_ISSET(int fd, fd\_set \*set);

FD\_SET(int fd, fd\_set \*set);

FD\_ZERO(int fd);

Функция select работает с тремя множествами дескрипторов, каждое из которых имеет тип fd\_set. В множество readfds записываются дескрипторы сокетов, из которых нам требуется читать данные (слушающие сокеты добавляются в это же множество). Множество writefds должно содержать дескрипторы сокетов, в которые мы собираемся писать, а exceptfds - дескрипторы сокетов, которые нужно контролировать на возникновение ошибки. Если какое-то множество вас не интересуют, вы можете передать вместо указателя на него NULL. Что касается других параметров, в n нужно записать максимальное значение дескриптора по всем множествам плюс единица, а в timeout - величину таймаута. Структура timeval имеет следующий формат.

struct timeval {

int tv\_sec; // секунды

int tv\_usec; // микросекунды

};

Поле "микросекунды" смотрится впечатляюще. Но на практике вам не добиться такой точности измерения времени при использовании select. Реальная точность окажется в районе 100 миллисекунд.

Теперь займёмся множествами дескрипторов. Для работы с ними предусмотрены функции FD\_XXX, показанные выше; их использование полностью скрывает от нас детали внутреннего устройства fd\_set. Рассмотрим их назначение.

FD\_ZERO(fd\_set \*set) - очищает множество set

FD\_SET(int fd, fd\_set \*set) - добавляет дескриптор fd в множество set

FD\_CLR(int fd, fd\_set \*set) - удаляет дескриптор fd из множества set

FD\_ISSET(int fd, fd\_set \*set) - проверяет, содержится ли дескриптор fd в множестве set

Если хотя бы один сокет готов к выполнению заданной операции, select возвращает ненулевое значение, а все дескрипторы, которые привели к "срабатыванию" функции, записываются в соответствующие множества. Это позволяет нам проанализировать содержащиеся в множествах дескрипторы и выполнить над ними необходимые действия. Если сработал таймаут, select возвращает ноль, а в случае ошибки -1. Расширенный код записывается в errno.

Программы, использующие неблокирующие сокеты вместе с select, получаются весьма запутанными. Если в случае с fork мы строим логику программы, как будто клиент всего один, здесь программа вынуждена отслеживать дескрипторы всех клиентов и работать с ними параллельно.