Michael Whittaker

Beej's Guide to Network Programming

Thoughts and Notes

beejsockets github

Contents

Listings

Preface

This document contains the notes, musings, and thoughts generated during my reading of "Beej's Guide to Network Programming" by Brian Hall. The notes were taken primarily to encourage a thorough reading of the book and to help me recall the most important tidbits from the book upon a rereading of my notes. I can imagine the notes may be helpful to more than just me, so I am making them publicly available. A network programming novice, I cannot guarantee my notes are entirely correct, or even sensical at times. If you ever encounter a mistake, please contact me at mjw297@cornell.edu.

Along with the notes, I've thrown together some source code and other resources. Some of the code is taken directly from the text while some is original. All notes, code, and resources can be found at the beejsockets github.

Enjoy!

1 Intro

This book will teach network programming!

1.1 Audience

This is a tutorial, not a reference, for novice programmers.

1.2 Platform and Compiler

Compiled using Gnu's gcc.

1.3 Official Homepage and Books For Sale

Visit http://beej.us/guide/bgnet and http://beej.us/guide/url/bgbuy.

1.4 Note for Solaris/SunOS Programmers

You have to additional work (see the book).

1.5 Note for Windows Programmers

Switch to Unix :P

2 What is a socket?

A *socket* is a way to speak to other programs using standard Unix file descriptors. Recall that everything in Unix is a file. All I/O is done by reading and writing to a file descriptor, an integer associated with an open file. The file, however, can be many things: a network connection, a FIFO, a pipe, a terminal, etc. If we want to communicate with another program over the Internet, we'll do it via a file descriptor.We get, read, and write sockets using the `socket()`, `send()`, and `recv()` system calls.

There are many different kinds of sockets. This book will deal with DARPA Internet sockets.

2.1 Two Types of Internet Sockets

There are two types of sockets: "Stream Sockets" and "Datagram Sockets", also known as `SOCK_STREAM` and `SOCK_DGRAM` respectively.

Stream sockets Stream sockets are reliable two-way connected communication streams. The order of sent messages are maintained and the messaging is guaranteed to be error-free.

Applications such as telnet and the HTTP protocol use stream sockets.

Stream sockets use the Transmission Control Protocol, TCP, to guarantee their reliability.

Datagram sockets Datagram sockets are unreliable and connectionless. If you send a message it may not arrive and it may not arrive in the correct order. The only guarantee is that if the message does arrive, the data inside will be error-free.

Datagram sockets are connectionless because unlike stream sockets, you don't have to maintain an open connection. They are typically used in applications where dropping a few packets here and there is not important.

tftp, dhcpcd, multiplayer games, audio streaming, and video conferencing, all can use datagram sockets. Some applications like tftp and dhcpcd need additional protocols on top of UDP to ensure the packets make it, but other applications like gaming will simply ignore dropped packets (e.g. lag).

You would use and unreliable protocol like UDP for speed!

2.2 Low level Nonsense and Network Theory

Data encapsulation is how networking works. Essentially, data is wrapped in various headers and sent out, such as in Figure 1. When the packet is received, hardware will strip the ethernet header. The kernel will strip the IP and UDP headers. A TFTP program will stip the TFTP header and manipulate the unencapsulated data.

Such encapsulation is used in the *Layered Network Model*.

- Application

Figure 1: Data Encapsulation

- Presentation
- Session
- Transport
- Network
- Data Link
- Physical

A model more consistent with Unix might be:

- Application Layer (telnet, ftp, etc.)
- Host-to-Host Transport Layer (TCP, UDP)
- Internet Layer (IP and routing)
- Network Access Layer (Ethernet, wi-fi, etc.)

3 IP Adresses, structs, and Data Munging

This section discusses IP addresses and ports as well as how the sockets API stores and manipulates IP addresses and other data.

3.1 IP Addresses, versions 4 and 6

Back when the Internet was originally created, we used IPv4. IP addresses were 32 bits (4 octets) and represented with "dots and numbers" as in `192.0.2.111`. However, as the number of required IP addresses grew, we ran out of IP addresses.

Enter IPv6. IPv6 addresses are 16 octets long and represented with colons and hexadecimal, as in `2001:odb8:c9d2:aee5:73e3:934a:a5ae:9551`. To compress IPv6 addresses, you can replace zeros with to colons. You can also leave off leading zeros in each byte pair. All of the pairs of IP addresses in Listing 1 are identical.

Listing 1: Identical IP Addresses

```
2001:0db8:c9d2:0012:0000:0000:0000:0051
2001:db8:c9d2:12::51

2001:0db8:ab00:0000:0000:0000:0000:0000
2001:db8:ab00::

0000:0000:0000:0000:0000:0000:0000:0001
::1
```

`::1` is the *loopback address* which is `127.0.0.1` in IPv4. IPv4 addresses can also be represented in IPv6. `192.0.2.33` translates to `::ffff:192.0.2.33`.

3.1.1 Subnets

For organizational purposes, it is convenient to label the first part of an IP address as the *network portion* of the address and the remaining part as the *host portion*. For example, consider the address `192.0.2.12`. The first three bytes could be the network and the last byte could be the host. That is, host `12` on network `192.0.2.0`.

In early versions of the Internet, there were different "classes" of subnets. Class A subnets had 1 byte of network. Class B subnets had 2 bytes of network. Class C subnets had 3 bytes of network. Eventually this scheme was deprecated and replaced with arbitrary length network portions.

The network portion of an address is described by a *netmask*, a set of bits you bitwise-AND the address with. For example, the netmask `255.255.255.0` yields three bytes of network. This scheme can also be expressed as an address followed by a forward slash followed by the number of bits in the network portion of the address. For example, `192.0.2.12/30` or `2001:db8::/32`.

3.1.2 Port Numbers

How do you multiplex different TCP or UDP applications on a computer with a single IP address? You use port numbers, a 16-bit number. Think of IP addresses as hotel addresses and port numbers as room numbers. Different applications run on different port numbers. HTTP runs on port 80, telnet on port 23, DOOM on port 666, etc.

3.2 Byte Order

Pretend you want to store the bytes `b34f` in your computer. Your computer can store them as `b3` then `4f`. This method, with the big end first, is known as *Big-Endian*. Other computers may store the bytes as `4f` then `b3` in a method known as *Little-Endian*.

Network Byte Order is synonymous with Big-Endian, and is the byte ordering sent across the network. *Host Byte Order* is the byte ordering of your computer. To convert to and from host and network ordering, we use 4 functions.

- `htons` host to network short
- `htonl` host to network long
- `ntohs` network to host short
- `ntohl` network to host long

3.3 structs

Refer to Listing 2, Listing 3, Listing 4, Listing 5, Listing 6, Listing 7, and Listing 8.

A socket descriptor is of type `int`.

A `addrinfo` struct is one of the first structs you'll interact with. It contains information about an address.

Listing 2: addrinfo struct

```
struct addrinfo {
    int              ai_flags;     // AI_PASSIVE, AI_CANONNAME, etc.
    int              ai_family;    // AF_INET, AF_INET6, AF_UNSPEC
    int              ai_socktype;  // SOCK_STREAM, SOCK_DGRAM
    int              ai_protocol;  // use 0 for "any"
    size_t           ai_addrlen;   // size of ai_addr in bytes
    struct sockaddr *ai_addr;      // struct sockaddr_in or _in6
    char            *ai_canonname; // full canonical hostname

    struct addrinfo *ai_next;      // linked list, next node
};
```

`sockaddr` contains a socket address. Dealing with the `sa_data` by hand is cumbersome. Instead, you can use `sockaddr_in` or `sockaddr_in6`. A pointer to a `sockaddr_in` can be cast to a pointer of `sockaddr` and vice-versa.

Listing 3: sockaddr struct

```
struct sockaddr {
    unsigned short sa_family;   // address family, AF_xxx
    char           sa_data[14]; // 14 bytes of protocol address
};
```

Listing 4: sockaddr-in struct

```
struct sockaddr_in {
    short int          sin_family;  // Address family, AF_INET
    unsigned short int sin_port;    // Port number
    struct in_addr     sin_addr;    // Internet address
    unsigned char      sin_zero[8]; // Same size as struct sockaddr
};
```

Listing 5: in-addr struct

```
struct in_addr {
    uint32_t s_addr; // that's a 32-bit int (4 bytes)
};
```

Listing 6: sockaddr-in6 struct

```
struct sockaddr_in6 {
    u_int16_t       sin6_family;   // address family, AF_INET6
    u_int16_t       sin6_port;     // port number, Network Byte Order
    u_int32_t       sin6_flowinfo; // IPv6 flow information
    struct in6_addr sin6_addr;     // IPv6 address
    u_int32_t       sin6_scope_id; // Scope ID
};
```

Listing 7: in6-addr struct

```
struct in6_addr {
    unsigned char s6_addr[16]; // IPv6 address
};
```

`sockaddr_storage` is a struct large enough to hold both IPv4 and IPv6 structures.

Listing 8: sockaddr-storage struct

```
struct sockaddr_storage {
    sa_family_t ss_family; // address family

    // all of this is padding, implementation specific, ignore it:
    char      __ss_pad1[_SS_PAD1SIZE];
    int64_t   __ss_align;
    char      __ss_pad2[_SS_PAD2SIZE];
};
```

3.4 IP Addresses, Part Deux

Fortunately, there are many functions to help manipulate IP addresses.

If you want to convert a string representation of an IP address into a representation suitable for a struct, use `inet_pton()`. `pton` stands for "presentation to network" or "printable to network". An example use is given in Listing 9. `inet_pton` returns -1 on error and 0 if the address is messed up.

Listing 9: Presentation to Network

```
struct sockaddr_in  sa;  // IPv4
struct sockaddr_in6 sa6; // IPv6

inet_pton(AF_INET, "192.0.2.1", &(sa.sin_addr));
inet_pton(AF_INET6, "2001:db8:63b3:1::3490", &(sa6.sin6_addr));
```

The opposite conversion can be made using `inet_ntop()`, as shown in Listing 10.

Listing 10: Network to Presentation

```
char ip4[INET_ADDRSTRLEN];
struct sockaddr_in sa;
inet_ntop(AF_INET, &(sa.sin_addr), ip4, INET_ADDRSTRLEN);
printf("The IPv4 address is: %s\n", ip4);

char ip6[INET6_ADDRSTRLEN];
struct sockaddr_in6 sa6;
inet_ntop(AF_INET6, &(sa6.sin6_addr), ip6, INET6_ADDRSTRLEN);
printf("The address is: %s\n", ip6);
```

3.4.1 Private (Or Disconnected) Networks

Many networks are hidden behind a firewall that translates internal IP addresses to external IP addresses. This is done via a process known as NAT, or *Network Address Translation*.

Your public IP may be `192.0.2.33`, but your computer will say `10.x.x.x` or `192.168.x.x`.

4 Jumping from IPv4 to IPv6

5 System Calls or Bust

6 Client-Server Background

7 Slightly Advanced Techniques