# Network Programming

## Lecture 5—Advanced Sockets I: Raw Sockets and Datalink Access

Lei Wang
lei.wang@dlut.edu.cn

Dalian University of Technology

Dec 15, 2008

# Part 3. Advanced Sockets I: Raw Sockets and Datalink Access

1. Raw Socket
   - Introduction
   - Raw Socket Creation, Output and Input
   - `ping` and `traceroute` Program
2. Datalink Access
   - Introduction
   - BPF, DLPI and Linux: `SOCK_PACKET` and `PF_PACKET`
   - `libpcap`: Packet Capture Library
   - `libnet`: Packet Creation and Injection Library
   - Example: Examining the UPD Checksum Field

# Introduction

Raw sockets provide three features not provided by normal TCP and UDP sockets:

- Raw sockets let us read and write ICMPv4, IGMPv4, and ICMPv6 packets. (`ping`, `mrouted`)
- With a raw socket, a process can read and write IPv4 datagrams with an IPV4 protocol field that is not processed by the kernel.
- With a raw socket, a process can build its own IPv4 header using the `IP_HDRINCL` socket option.

# Raw Socket Creation

- SOCKET_RAW

```
int     sockfd;

sockfd = socket(AF_INET, SOCK_RAW, protocol);
```

  where protocol is one of the constants, IPPROTO_xxx, such as IPPROTO_ICMP.

- IP_HDRINCL socket option
- bind can be called on the raw socket, but this is rare.
- connect can be called on the raw socket, but this is rare.

# Raw Socket Output

- Normal output is performed by calling `sendto` or `sendmsg` and specifying the destination IP address.
- If the `IP_HDRINCL` option is ***not*** set, the starting address of the data for the kernel to send specifies the first byte following the IP header because the kernel will build the IP header and prepend it to the data from the process.
- If the `IP_HDRINCL` option is set, the starting address of the data for the kernel to send specifies the first byte of the IP header.
- The kernel fragments raw packets that exceed the outgoing interface MTU.

# Raw Socket Input

Which received IP datagrams does the kernel pass to raw sockets? The following rules apply:

- Received UDP packets and received TCP packets are never passed to a raw socket. (must be read at the datalink layer)
- Most ICMP packets are passed to a raw socket after the kernel has finished processing the ICMP message.
- All IGMP packets are passed to a raw socket after the kernel has finished processing the IGMP message.
- All IP datagrams with a protocol field that the kernel does not understand are passed to a raw socket.
- If the datagram arrives in fragments, nothing is passed to a raw socket until all fragments have arrived and have been reassembled.

# Raw Socket Input Contd.

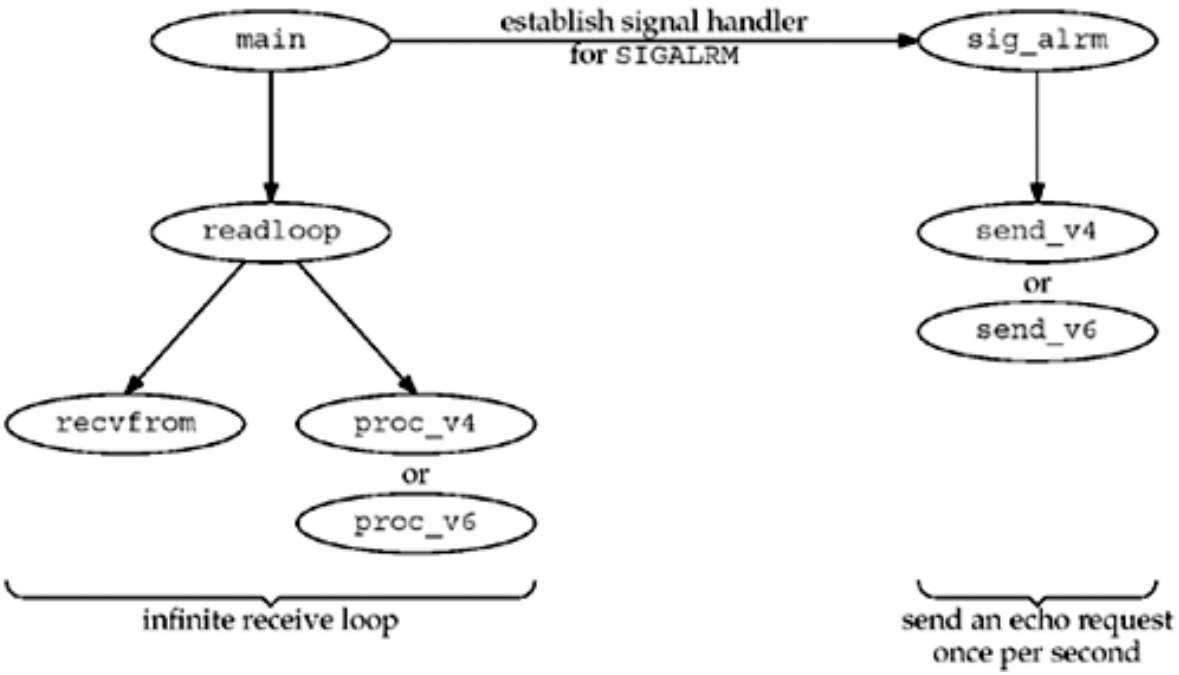
When the kernel has an IP datagram to pass to the raw sockets, all raw sockets for all processes are examined, looking for all matching sockets. A copy of the IP datagram is delivered to each matching socket. The following tests are performed for each raw socket and only if all three tests are true is the datagram delivered to the socket:

- If a nonzero protocol is specified when the raw socket is created (the third argument to `socket`), then the received datagram's protocol field must match this value or the datagram is not delivered to this socket.
- If a local IP address is bound to the raw socket by `bind`, then the destination IP address of the received datagram must match this bound address or the datagram is not delivered to this socket.
- If a foreign IP address was specified for the raw socket by `connect`, then the source IP address of the received datagram must match this connected address or the datagram is not delivered to this socket.

Notice that if a raw socket is created with a protocol of 0, and neither `bind` nor `connect` is called, then that socket receives a copy of every raw datagram the kernel passes to raw sockets.

# ping Program

- A ping program without suffering *creeping featurism*.
- Overview of the functions in the ping program.



# traceroute Program

- traceroute lets us determine the path that IP datagrams follow from our host to other destination.
- Exploit the IPv4 TTL field, ICMP "time exceeded in transit" and ICMP "port unreachable" error.
- IP_HDRINCL and IP_TTL

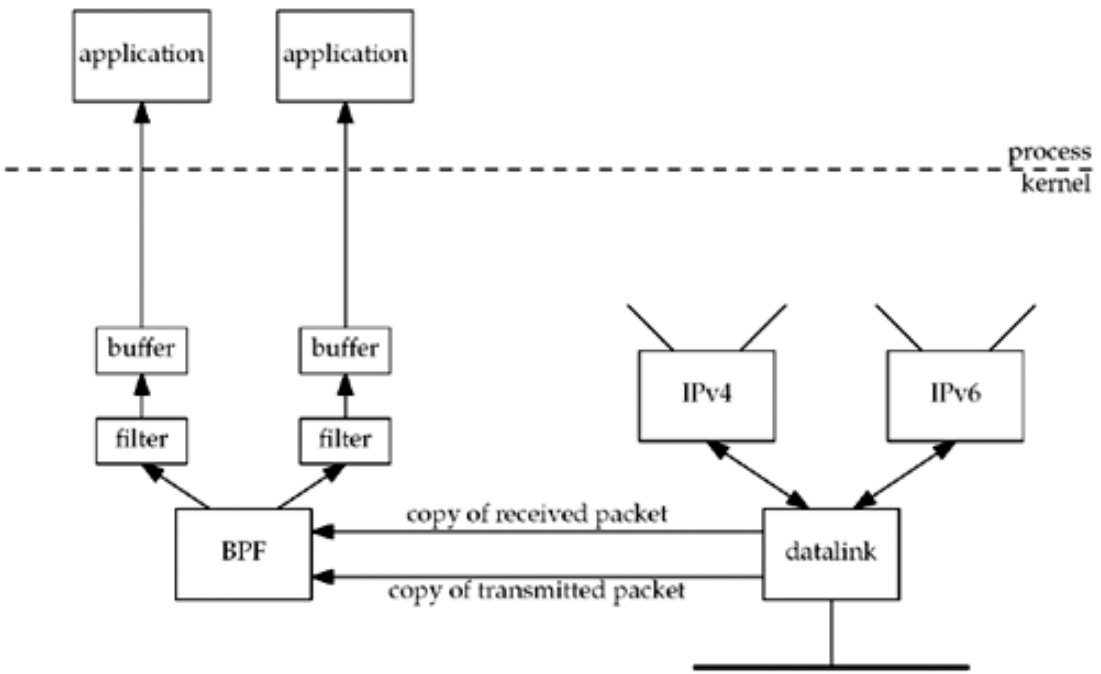
# Datalink Access

Providing access to the datalink layer for an application, with the following capabilities:

- The ability to watch the packets received by the datalink layer, allowing programs such as `tcpdump` to be run on normal computer systems (as opposed to dedicated hardware devices to watch packets). (***promiscuous mode***)
- The ability to run certain programs as normal applications instead of as part of the kernel. For example, most Unix versions of an RARP server are normal applications that read RARP requests from the datalink (RARP requests are not IP datagrams) and then write the reply back to the datalink.
- Three common methods under Unix are: the BSD Packet Filter (BPF), the SVR4 Datalink Provider Interface (DLPI), and the Linux `SOCK_PACKET` interface.

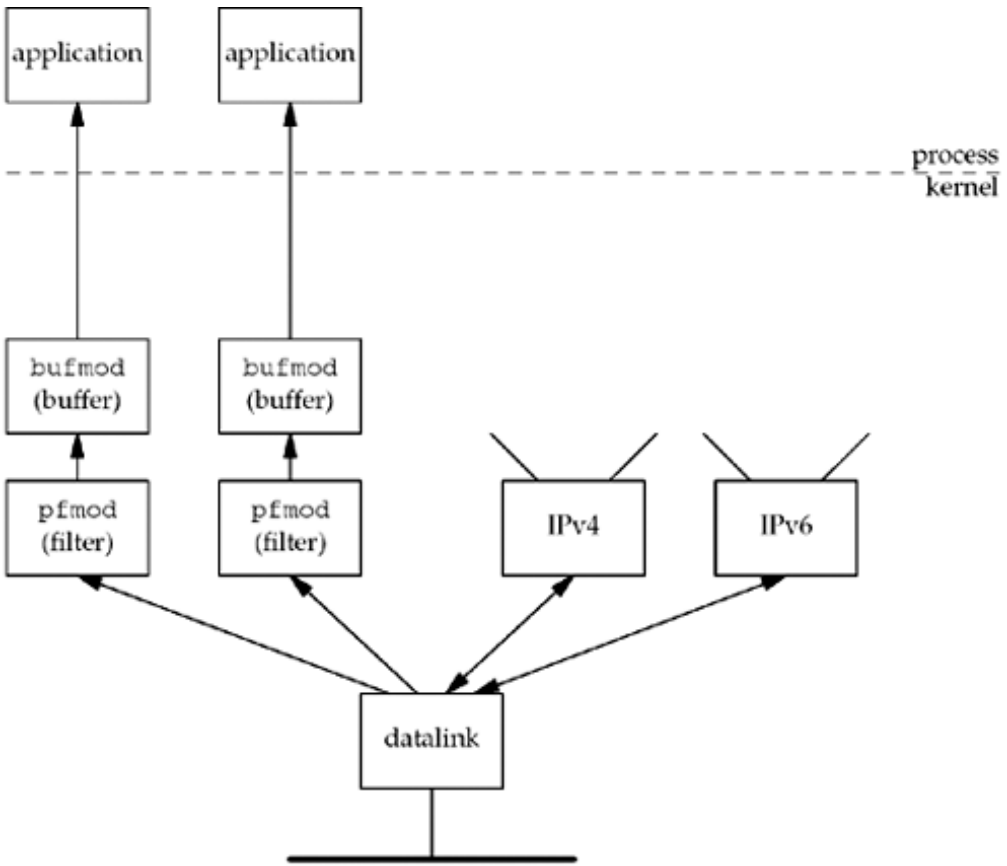
# BSD Packet Filter (BPF)

4.4BSD and many other Berkeley-derived implementations support BPF.



# Datalink Provide Interface (DLPI)

SVR4 provides datalink access through DLPI.



# Linux: SOCK_PACKET and PF_PACKET

Two methods under Linux:

- SOCK_PACKET: original, widely available but less flexible
- PF_PACKET: newer method

```
/* newer systems*/
fd = socket(PF_PACKET, SOCK_RAW, htons(ETH_P_ALL));

/* older systems*/
fd = socket(AF_INET, SOCK_PACKET, htons(ETH_P_ALL));
```

# libpcap: Packet Capture Library

- libpcap provides implementation-independent access to the underlying packet capture facility provided by the OS.
- Currently, it supports only the reading of packets (although adding a few lines of code to the library lets one write datalink packets too on some systems).
- The library is available from http://www.tcpdump.org/.

# libnet: Packet Creation and Injection Library

- libnet provides an interface to craft and inject arbitrary packets into the network.
- It provides both raw socket and datalink access modes in an implementation-independent manner.
- http://www.packetfactory.net/libnet/

# Example: Examining the UPD Checksum Field