Network stack in Kernel - Presentation

Tom Decrette

The network stack

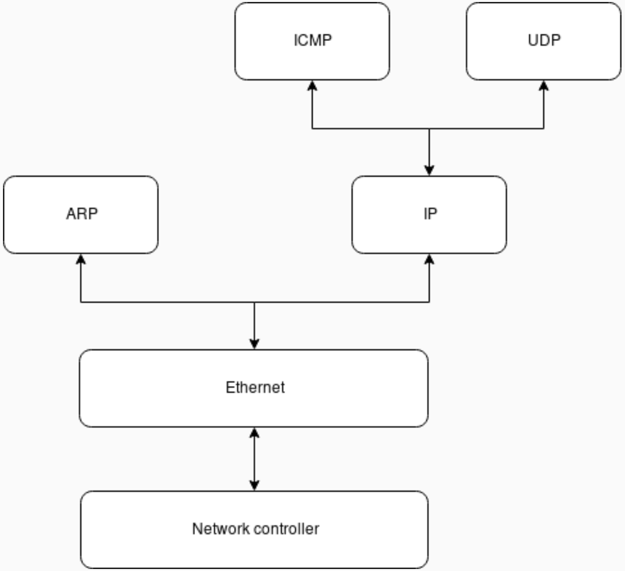


Figure 1: Network Stack

Interactions through the stack

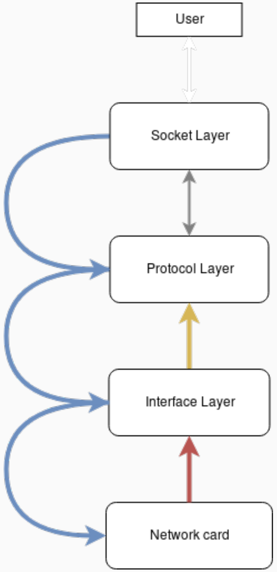


Figure 2: Interactions

Memory Buffer

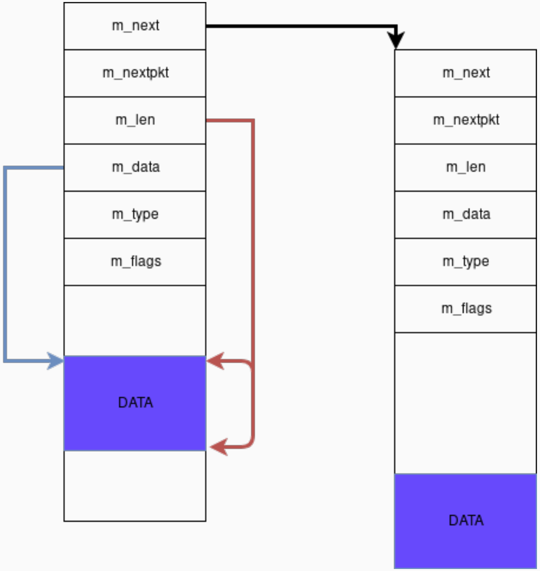


Figure 3: Mbuf chained

Memory Buffer

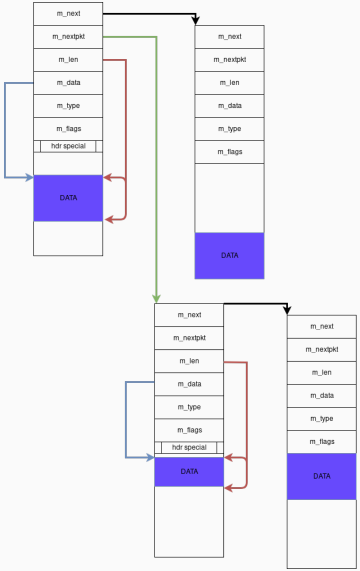
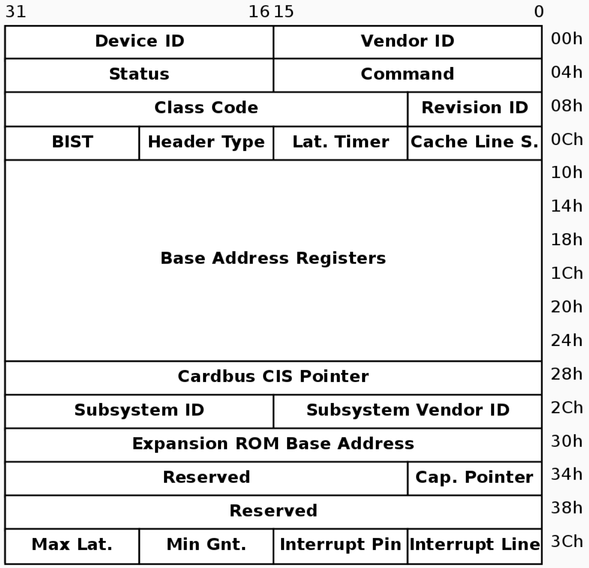


Figure 4: Mbuf chained headers

What is PCI?



31	16	15	0	
Device ID		Vendor ID		00h
Status		Command		04h
Class Code			Revision ID	08h
BIST	Header Type	Lat. Timer	Cache Line S.	0Ch
Base Address Registers				10h
				14h
				18h
				1Ch
				20h
				24h
Cardbus CIS Pointer				28h
Subsystem ID		Subsystem Vendor ID		2Ch
Expansion ROM Base Address				30h
Reserved			Cap. Pointer	34h
Reserved				38h
Max Lat.	Min Gnt.	Interrupt Pin	Interrupt Line	3Ch

Figure 5: PCI Header

Finding a device

```
for(bus = 0; bus < 256; bus++) {
    for(device = 0; device < 32; device++) {
        if(vendorID == 0xFFFF)
            continue;        // Device doesn't exist
        if (is_device(bus, device, 0))
            //Device is found
            return;
        headerType = getHeaderType(bus, device, function);
        if((headerType & 0x80) != 0) {
         /* It is a multi-function device, so check remaining functions */
            for(function = 1; function < 8; function++) {
                if(getVendorID(bus, device, function) != 0xFFFF && is_device(bus, device, function))
                    return;
             }

         }
     }
 }
```

Initialisation of network card

Many registers to use:

Name	Abbreviation	Value to use
Device Control	CTRL	CTRL_SLU
Receive/Transmit	TCTL / RCTL	BSIZE, SECRC, EN, ...
Base Address	_DBAL / _DBAH	Queue Base Address
Queue Length	RDLEN / TDLEN	Number of descriptors
Queue Head	RDH / TDH	First desc. index
Queue Tail	RDT / TDT	Last desc. index
Interrupt Mask	IMS	RXO, RXT0, LSC, TXDE, ...

Network ring buffer

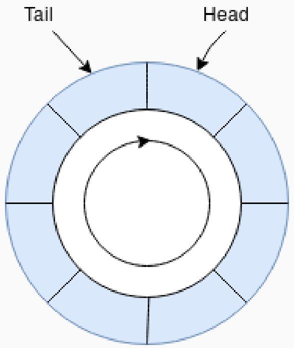


Figure 6: Ring Buffer Empty

Network ring buffer

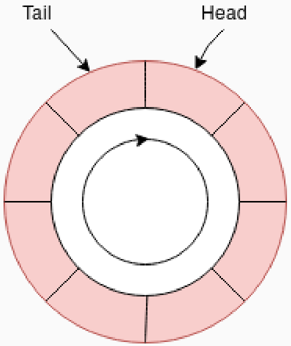


Figure 7: Ring Buffer Full

Descriptor

```c
struct e1000_descriptor {
    u64 addr;
    u16 length;
    union  {
        u16 checksum; /* RX */
        struct { /* TX */
            u8 cso;
            u8 cmd;
        };
    };
    u8 status;
    union {
        u8 errors; /* RX */
        u8 css; /* TX */
    };
    u16 special;
} __attribute__((packed));
```

Ethernet handling

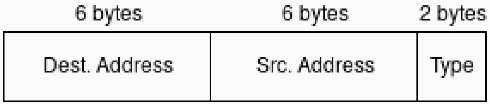


Figure 8: Ethernet Header

IP handling

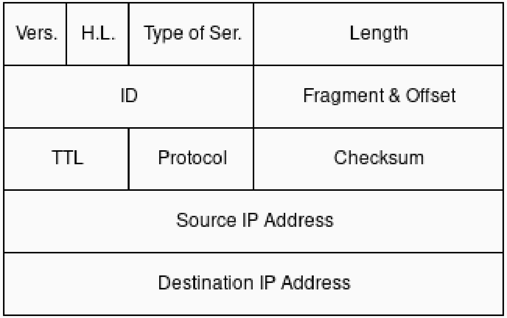


Figure 9: IP Header

ICMP handling

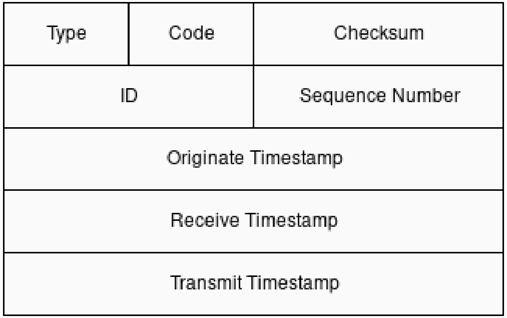
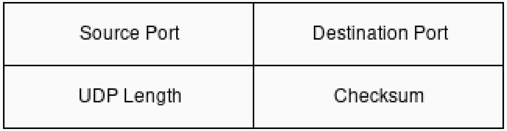


Figure 10: ICMP Header

UDP handling



Source Port	Destination Port
UDP Length	Checksum

Figure 11: UDP Header

Sockets

- Interface Kernel/Userland
- Store data to send and to receive
- Easily changed through syscalls
- One socket per distant IP

System calls handling

The only system calls that are important here are:

- bind
- gethost
- recvfrom
- sendto
- socket