CS61C
Anatomy of I/O Devices: Networks

Lecture 14

March 5, 1999

Dave Patterson
(http.cs.berkeley.edu/~patterson)

www-inst.eecs.berkeley.edu/~cs61c/schedule.html
Review 1/2

- Operating System started as shared I/O library
  - Support for OS abstraction: Kernel/User bit, stacked KU bits, syscall
  - MIPS follows coprocessor abstraction to add resources, instructions for OS

- Interrupt control: Interrupt Enable bit, stacked IE bits, Interrupt Priority Levels, Interrupt Mask
  - Re-entrant via restricting int. to higher priority

- DMA to accelerate data movement
Outline

° Buses
° Why Networks?
° A Simple Example: Derive Network Basics
° Protocol, Ethernet
° Administrivia, “Computers in the News”
° Internetworking, Protocol Suites, TCP/IP
° Performance Pitfalls
° Conclusion
Recall: 5 components of any Computer

- **Processor** (active)
  - Control (“brain”)
  - Datapath (“brawn”)

- **Memory** (passive)
  - (where programs, data live when running)

- **Devices**
  - Input
  - Output

- **Lectures 1-11**
- **Lectures 12-15**

- **Input Devices**: Keyboard, Mouse
- **Output Devices**: Disk, Network, Display, Printer
Connecting to Networks (and Other I/O)

° **Bus** - shared medium of communication that can connect to many devices

° Hierarchy of Buses in a PC
Buses in a PC

- **Data rates**
  - Memory: 100 MHz, 8 bytes ⇒ 800 MB/s (peak)
  - PCI: 33 MHz, 4 bytes wide ⇒ 132 MB/s (peak)
  - SCSI: “Ultra2” (40 MHz), “Wide” (2 bytes) ⇒ 80 MB/s (peak)
Why Networks?

- Originally sharing I/O devices between computers (e.g., printers)
- Then Communicating between computers (e.g., file transfer protocol)
- Then Communicating between people (e.g., email)
- Then Communicating between networks of computers ⇒ Internet, WWW
Types of Networks

° Local Area Network (Ethernet)
  • Inside a building: Up to 1 km
  • (peak) Data Rate:
    10 Mbits/sec, 100 Mbits/sec, 1000 Mbits/sec
  • Run, installed by network administrators

° Wide Area Network
  • Across a continent (10km to 10000 km)
  • (peak) Data Rate:
    1.5 Mbits/sec to 2500 Mbits/sec
  • Run, installed by telephone companies
ABCs of Networks: 2 Computers

° **Starting Point:** Send bits between 2 computers

° Queue (First In First Out) on each end

° Can send both ways ("Full Duplex")

° Information sent called a "message"
  • Note: Messages also called packets
A Simple Example: 2 Computers

What is Message Format?
- (Similar in idea to Instruction Format)
- Fixed size? Number bits?

Request/Response

<table>
<thead>
<tr>
<th>Address/Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 bit</td>
</tr>
<tr>
<td>32 bits</td>
</tr>
</tbody>
</table>

0: Please send data from address in your memory
1: Packet contains data corresponding to request

- **Header(Trailer)**: information to deliver message
- **Payload**: data in message (1 word above)
Questions About Simple Example

What if more than 2 computers want to communicate?

- Need computer “address field” in packet to know which computer should receive it (destination), and to which computer it came from for reply (source)

<table>
<thead>
<tr>
<th>Req./Resp.</th>
<th>Dest.</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net ID</td>
<td>Net ID</td>
<td>Address/Data</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1 bit</th>
<th>5 bits</th>
<th>5 bits</th>
<th>32 bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Header</td>
<td></td>
<td>Payload</td>
<td></td>
</tr>
</tbody>
</table>
Questions About Simple Example

What if message is garbled in transit?

- Add redundant information that is checked when message arrives to be sure it is OK
- 8-bit sum of other bytes: called “Check sum”; upon arrival compare check sum to sum of rest of information in message

<table>
<thead>
<tr>
<th>Req./Resp.</th>
<th>Dest.</th>
<th>Source</th>
<th>Checksum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Net ID</td>
<td>Net ID</td>
<td>Address/Data</td>
</tr>
</tbody>
</table>

1 bit 5 bits 5 bits 32 bits 8 bits

Header Payload Trailer
Questions About Simple Example

What if message never arrives?

- If tell sender it has arrived (and tell receiver reply has arrived), can resend upon failure
- Don’t discard message until get “ACK”;
  (Also, if check sum fails, don’t send ACK)

<table>
<thead>
<tr>
<th>Req./Resp.</th>
<th>Dest.</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Net ID</td>
<td>Net ID</td>
</tr>
<tr>
<td>2 bits</td>
<td>5 bits</td>
<td>5 bits</td>
</tr>
</tbody>
</table>

00: Request—Please send data from Address
01: Reply—Message contains data corresponding to request
10: Acknowledge (ACK) request
11: Acknowledge (ACK) reply
Observations About Simple Example

° Simple questions such as those above lead to more complex procedures to send/receive message and more complex message formats

° **Protocol**: algorithm for properly sending and receiving messages (packets)
Ethernet Packet Format

<table>
<thead>
<tr>
<th>Preamble</th>
<th>Dest Addr</th>
<th>Src Addr</th>
<th>Data</th>
<th>Pad</th>
<th>Check</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 Bytes</td>
<td>6 Bytes</td>
<td>6 Bytes</td>
<td>0-1500B</td>
<td>0-46B</td>
<td>4B</td>
</tr>
</tbody>
</table>

Length of Data
2 Bytes

- Preamble to recognize beginning of packet
- Unique Address per Ethernet Network Interface Card so can just plug in & use
- Pad ensures minimum packet is 64 bytes
  - Easier to find packet on the wire
- Header + Trailer: 24B + Pad
Software Protocol to Send and Receive

° SW Send steps

1: Application copies data to OS buffer
2: OS calculates checksum, starts timer
3: OS sends data to network interface HW and says start

° SW Receive steps

3: OS copies data from network interface HW to OS buffer
2: OS calculates checksum, if OK, send ACK; if not, delete message (sender resends when timer expires)
1: If OK, OS copies data to user address space, & signals application to continue
Administrivia 1/2

° 6th homework: Due 3/10 7PM
  • Exercises 8.3, 8.29 (skip challenge), Ap A.3

° 4th Project: Friday 3/12 7PM
  (absolute latest: 3/13 8AM)

° Readings: Cache Memory 7.1, 7.2

° Upcoming events
  • Midterm Review Sunday 3/14 2PM, 1 Pimentel
  • Midterm on Wed. 3/17 5pm-8PM, 1 Pimentel
  • No discussion sections 3/18, 3/19
  • Friday before Break 3/19: video tape by Gordon Moore, “Nanometers and Gigabucks”
Administrivia 2/2

° Copy of Slides for Midterm?
  • 150 pages total, 4/page double sided
  • Purchase at Copy Central
  • Is this worth doing?

° Printing Problems for lectures?
  • 20% don’t have access to printer at home
  • Could ask my assistant to make, say, 50 copies of slides back-to-back and place them in 271 Soda (or somewhere) by 10AM day of lecture
  • Is this worth doing?
“Computers in the News”


- “... an imminent privacy threat or simply part of the foundation of advanced computer systems”
- “Privacy groups argue fiercely that the merger of computers and the Internet has brought the specter of a new surveillance society (anything)”
- “...network designers argue that identity information is a vital aspect of modern security design because it is necessary to authenticate an individual in a network, thereby preventing fraud or intrusion”
- “all devices connected to networks require identification simply to function correctly... [for 20 years a] requirement for any computer connected to an Ethernet”
Protocol for Networks of Networks?

- **Internetworking**: allows computers on independent and incompatible networks to communicate reliably and efficiently;
  
  - Enabling technologies: SW standards that allow reliable communications without reliable networks
  
  - Hierarchy of SW layers, giving each layer responsibility for portion of overall communications task, called *protocol families* or *protocol suites*

- Abstraction to cope with **complexity of communication** vs. Abstraction for complexity of **computation**
Protocol for Network of Networks

Transmission Control Protocol/Internet Protocol (TCP/IP)

- This protocol family is the basis of the Internet, a WAN protocol
- IP makes best effort to deliver
- TCP guarantees delivery
- TCP/IP so popular it is used even when communicating locally: even across homogeneous LAN
FTP From Stanford to Berkeley

- BARRNet is WAN for Bay Area
  - T3 is 45 Mbit/s leased line (WAN);
    FDDI is 100 Mbit/s LAN

- IP sets up connection, TCP sends file
Protocol Family Concept

[Diagram showing the relationship between actual and logical messages]

Actual → Logical → Actual

Actual
Protocol Family Concept

° Key to protocol families is that communication occurs logically at the same level of the protocol, called peer-to-peer, but is implemented via services at the lower level

• Danger is each level lower performance if family is implemented as hierarchy (e.g., multiple check sums)
TCP/IP packet, Ethernet packet, protocols

- Application sends message
- TCP breaks into 64KB segments, adds 20B header
- IP adds 20B header, sends to network
- If Ethernet, broken into 1500B packets with headers, trailers (24B)
- All Headers, trailers have length field, destination, ...
**Shared vs. Switched Based Networks**

*Shared Media vs. Switched: pairs communicate at same time: “point-to-point” connections*

*Aggregate BW in switched network is many times shared*

- point-to-point faster since no arbitration, simpler interface
Example of Network Performance

° Compare Ethernet, ATM sending a message with a 5 MB payload

° Time to send message: Overhead + Size/Data Rate

° Ethernet Overhead: 500 microseconds
  • BW is 1.25 Mbyte/sec

° ATM Overhead: 600 microseconds
  • BW is 10 Mbyte/sec

° Transmission time = Overhead + Size/(Data Rate)
Example of Network Performance

° Compare Ethernet (10 Mbit/sec), ATM (800 Mbit/sec) sending a message with a 250 Byte payload

° Time to send message:
  Overhead + Size/Data Rate

° Ethernet Overhead: 460 microseconds
  • BW is 1.25 Mbyte/sec
  • Time = 460 + 250/1.25 = 660 usecs

° ATM Overhead: 630 microseconds
  • BW is 10 Mbyte/sec
  • Time = 630 + 250/10 = 655 microseconds
Limits to Performance of Networks

- Layers of protocol
- Processor speed for protocol processing
- Memory speed for transferring messages
- Bus speed for connecting to computer
- Sharing of Network by multiple computers
I/O Pitfall: Relying on Peak Data Rates

- Using the peak transfer rate of a portion of the I/O system to make performance projections or performance comparisons

- Peak bandwidth measurements often based on unrealistic assumptions about system or unattainable because of other system limitations

  - In example, Peak Bandwidth FDDI vs. 10 Mbit Ethernet = 10:1, but delivered BW ratio (due to software overhead) is 1.01:1

  - Peak PCI BW is 132 MByte/sec, but combined with memory often < 80 MB/s
Network Media (if time)

Twisted Pair: Copper, 1mm thick, twisted to avoid antenna effect (telephone)

Coaxial Cable:
- Copper core
- Insulator
- Braided outer conductor
- Plastic covering

Used by cable companies: high BW, good noise immunity

Fiber Optics
- Total internal reflection
- Transmitter
  - L.E.D
  - Laser Diode
- Light source
- Receiver
  - Photodiode
- Silica
- Air

Light:
- 3 parts are cable, light source, light detector
Protocol suites allow heterogeneous networking
• Another use of principle of abstraction
• Protocols ⇒ operation in presence of failures
• Standardization key for LAN, WAN

Integrated circuit revolutionizing network switches as well as processors
• Switch just a specialized computer

High bandwidth networks with slow SW overheads don’t deliver their promise

Next: Anatomy of disks, RAID