Chapter 5: The Data Link Layer

Our goals:

- understand principles behind data link layer services:
  - error detection, correction
  - sharing a broadcast channel: multiple access
  - link layer addressing
  - reliable data transfer, flow control: done!

- instantiation and implementation of various link layer technologies
Chapter 5 outline

- 5.1 Introduction and services
- 5.2 Error detection and correction
- 5.3 Multiple access protocols
- 5.4 LAN addresses and ARP
- 5.5 Ethernet
- 5.6 Hubs, bridges, and switches
- 5.7 Wireless links and LANs
- 5.8 PPP
- 5.9 ATM
- 5.10 Frame Relay
Link Layer: Introduction

Some terminology:

- hosts and routers are **nodes** (bridges and switches too)
- communication channels that connect adjacent nodes along communication path are **links**
  - wired links
  - wireless links
  - LANs
- 2-PDU is a **frame**, encapsulates datagram

**data-link layer** has responsibility of transferring datagram from one node to adjacent node over a link
Link layer: context

- Datagram transferred by different link protocols over different links:
  - e.g., Ethernet on first link, frame relay on intermediate links, 802.11 on last link
- Each link protocol provides different services
  - e.g., may or may not provide rdt over link

Transportation analogy
- trip from Princeton to Lausanne
  - limo: Princeton to JFK
  - plane: JFK to Geneva
  - train: Geneva to Lausanne
- tourist = datagram
- transport segment = communication link
- transportation mode = link layer protocol
- travel agent = routing algorithm
Link Layer Services

- **Framing, link access:**
  - Encapsulate datagram into frame, adding header, trailer
  - Channel access if shared medium
  - 'Physical addresses' used in frame headers to identify source, dest
    - Different from IP address!

- **Reliable delivery between adjacent nodes**
  - We learned how to do this already (chapter 3)!
  - Seldom used on low bit error link (fiber, some twisted pair)
  - Wireless links: high error rates
    - Q: Why both link-level and end-end reliability?
Link Layer Services (more)

- **Flow Control:**
  - pacing between adjacent sending and receiving nodes

- **Error Detection:**
  - errors caused by signal attenuation, noise.
  - receiver detects presence of errors:
    - signals sender for retransmission or drops frame

- **Error Correction:**
  - receiver identifies and corrects bit error(s) without resorting to retransmission

- **Half-duplex and full-duplex**
  - with half duplex, nodes at both ends of link can transmit, but not at same time
Adaptors Communicating

- link layer implemented in “adaptor” (aka NIC)
  - Ethernet card, PCMCI card, 802.11 card
- sending side:
  - encapsulates datagram in a frame
  - adds error checking bits, rdt, flow control, etc.
- receiving side
  - looks for errors, rdt, flow control, etc
  - extracts datagram, passes to rcving node
- adapter is semi-autonomous
- link & physical layers
Chapter 5 outline

- 5.1 Introduction and services
- 5.2 Error detection and correction
- 5.3 Multiple access protocols
- 5.4 LAN addresses and ARP
- 5.5 Ethernet
- 5.6 Hubs, bridges, and switches
- 5.7 Wireless links and LANs
- 5.8 PPP
- 5.9 ATM
- 5.10 Frame Relay
Error Detection

EDC = Error Detection and Correction bits (redundancy)

D = Data protected by error checking, may include header fields

• Error detection not 100% reliable!
  • protocol may miss some errors, but rarely
  • larger EDC field yields better detection and correction
Parity Checking

Single Bit Parity:
Detect single bit errors

\[ \begin{array}{cccccc}
0 & 1 & 1 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 1
\end{array} \]

Two Dimensional Bit Parity:
Detect and correct single bit errors

\[ \begin{array}{cccc}
\text{row parity} & d_1, j & \ldots & d_1, j+1 \\
\text{column parity} & d_2, 1 & \ldots & d_2, j+1 \\
& \ldots & \ldots & \ldots & \ldots \\
& d_{i, 1} & \ldots & d_{i, j} \\
& d_{i, j+1} & \ldots & d_{i+1, j} \\
& \ldots & \ldots & \ldots & \ldots \\
& d_{i+1, 1} & \ldots & d_{i+1, j+1}
\end{array} \]

0 errors

\[ \begin{array}{cccc}
1 & 0 & 1 & 0 \\
1 & 1 & 1 & 0 \\
0 & 1 & 1 & 1 \\
0 & 1 & 0 & 1
\end{array} \]

Correctable single bit error

\[ \begin{array}{cccc}
1 & 0 & 1 & 0 \\
0 & 1 & 1 & 0 \\
0 & 1 & 1 & 0 \\
0 & 1 & 0 & 1
\end{array} \]

Parity error
Internet checksum

**Goal:** detect “errors” (e.g., flipped bits) in transmitted segment (note: used at transport layer only)

**Sender:**
- treat segment contents as sequence of 16-bit integers
- checksum: addition (1’s complement sum) of segment contents
- sender puts checksum value into UDP checksum field

**Receiver:**
- compute checksum of received segment
- check if computed checksum equals checksum field value:
  - NO - error detected
  - YES - no error detected. But maybe errors nonetheless? More later ....
Checksumming: Cyclic Redundancy Check

- view data bits, $D$, as a binary number
- choose $r+1$ bit pattern (generator), $G$
- goal: choose $r$ CRC bits, $R$, such that
  - $\langle D,R \rangle$ exactly divisible by $G$ (modulo 2)
  - receiver knows $G$, divides $\langle D,R \rangle$ by $G$. If non-zero remainder: error detected!
  - can detect all burst errors less than $r+1$ bits
- widely used in practice (ATM, HDCL)

\[ D \ast 2^r \text{ XOR } R \]

$D$: data bits to be sent $R$: CRC bits

**bit pattern**

**mathematical formula**
CRC Example

Want:

\[ D \cdot 2^r \text{ XOR } R = nG \]
equivalently:

\[ D \cdot 2^r = nG \text{ XOR } R \]
equivalently:

if we divide \( D \cdot 2^r \) by \( G \),
want remainder \( R \)

\[ R = \text{remainder}[\frac{D \cdot 2^r}{G}] \]
Chapter 5 outline

- 5.1 Introduction and services
- 5.2 Error detection and correction
- 5.3 Multiple access protocols
- 5.4 LAN addresses and ARP
- 5.5 Ethernet
- 5.6 Hubs, bridges, and switches
- 5.7 Wireless links and LANs
- 5.8 PPP
- 5.9 ATM
- 5.10 Frame Relay
Multiple Access Links and Protocols

Two types of “links”:

- **point-to-point**
  - PPP for dial-up access
  - point-to-point link between Ethernet switch and host

- **broadcast** (shared wire or medium)
  - traditional Ethernet
  - upstream HFC
  - 802.11 wireless LAN
Multiple Access protocols

- single shared broadcast channel
- two or more simultaneous transmissions by nodes:
  - interference
    - only one node can send **successfully** at a time

**multiple access protocol**

- distributed algorithm that determines how nodes share channel, i.e., determine when node can transmit
- communication about channel sharing must use channel itself!
- what to look for in multiple access protocols:
Ideal Multiple Access Protocol

Broadcast channel of rate $R$ bps

1. When one node wants to transmit, it can send at rate $R$.

2. When $M$ nodes want to transmit, each can send at average rate $R/M$.

3. Fully decentralized:
   - no special node to coordinate transmissions
   - no synchronization of clocks, slots

4. Simple
MAC Protocols: a taxonomy

Three broad classes:

- **Channel Partitioning**
  - divide channel into smaller “pieces” (time slots, frequency, code)
  - allocate piece to node for exclusive use

- **Random Access**
  - channel not divided, allow collisions
  - “recover” from collisions

- **“Taking turns”**
  - tightly coordinate shared access to avoid collisions
Channel Partitioning MAC protocols: TDMA

TDMA: time division multiple access
- access to channel in "rounds"
- each station gets fixed length slot (length = pkt trans time) in each round
- unused slots go idle
- example: 6-station LAN, 1,3,4 have pkt, slots 2,5,6 idle

![Diagram of TDMA frame with stations 1, 3, 4, and 1, 3, 4 broadcasting at different times.]
Channel Partitioning MAC protocols: FDMA

FDMA: frequency division multiple access
▶ channel spectrum divided into frequency bands
▶ each station assigned fixed frequency band
▶ unused transmission time in frequency bands go idle
▶ example: 6-station LAN, 1,3,4 have pkt, frequency bands 2,5,6 idle
Channel Partitioning (CDMA)

CDMA (Code Division Multiple Access)

- unique “code” assigned to each user; i.e., code set partitioning
- used mostly in wireless broadcast channels (cellular, satellite, etc)
- all users share same frequency, but each user has own “chipping” sequence (i.e., code) to encode data
- encoded signal = (original data) X (chipping sequence)
- decoding: inner-product of encoded signal and chipping sequence
- allows multiple users to “coexist“ and transmit simultaneously with minimal interference (if codes are “orthogonal”)

Channel Partitioning (CDMA)
CDMA Encode/Decode

**Sender**
- Data bits: 
  - $d_1 = -1$
  - $d_0 = 1$
- Code:
  - Slot 1: 1 1 1 1
  - Slot 0: -1 -1 -1 -1

**Channel Output**
- $Z_{i,m} = d_i \cdot c_{m_i}$
- Slot 1 output: 1 1 1 1
- Slot 0 output: -1 -1 -1 -1

**Receiver**
- $d_i = \frac{\sum_{m=1}^{M} Z_{i,m} \cdot c_m}{M}$
- Code:
  - Slot 1 received input: 1 1 1 1
  - Slot 0 received input: -1 -1 -1 -1
- Output: $d_1 = -1$
  - $d_0 = 1$
CDMA: two-sender interference

senders

data bits

code

data bits

code

channel, $Z_{i,m}^*$

receiver 1

$Z_{i,m}^1 = d_{i}^1 \cdot c_{m}^1$

$Z_{i,m}^2 = d_{i}^2 \cdot c_{m}^2$

$d_i^1 = \sum_{m=1}^{M} Z_{i,m}^* \cdot c_{m}^1$

$M$

$d_i^0 = 1$

slot 1
received input

slot 0
received input
Random Access Protocols

- When node has packet to send
  - transmit at full channel data rate $R$
  - no a priori coordination among nodes
- two or more transmitting nodes $\rightarrow$ “collision”
- random access MAC protocol specifies:
  - how to detect collisions
  - how to recover from collisions (e.g., via delayed retransmissions)
- Examples of random access MAC protocols:
  - slotted ALOHA
  - ALOHA
  - CSMA, CSMA/CD, CSMA/CA
Slotted ALOHA

Assumptions
- all frames same size
- time is divided into equal size slots, time to transmit 1 frame
- nodes start to transmit frames only at beginning of slots
- nodes are synchronized
- if 2 or more nodes transmit in slot, all nodes detect collision

Operation
- when node obtains fresh frame, it transmits in next slot
- no collision, node can send new frame in next slot
- if collision, node retransmits frame in each subsequent slot with prob. p until success
Slotted ALOHA

Pros
- single active node can continuously transmit at full rate of channel
- highly decentralized: only slots in nodes need to be in sync
- simple

Cons
- collisions, wasting slots
- idle slots
- nodes may be able to detect collision in less than time to transmit packet
Slotted Aloha efficiency

Efficiency is the long-run fraction of successful slots when there’s many nodes, each with many frames to send.

- Suppose N nodes with many frames to send, each transmits in slot with probability p.
- prob that 1st node has success in a slot = \( p(1-p)^{N-1} \)
- prob that any node has a success = \( Np(1-p)^{N-1} \)

- For max efficiency with N nodes, find \( p^* \) that maximizes \( Np(1-p)^{N-1} \).
- For many nodes, take limit of \( Np^*(1-p^*)^{N-1} \) as N goes to infinity, gives \( 1/e = 0.37 \).

At best: channel used for useful transmissions 37% of time!
Pure (unslotted) ALOHA

- unslotted Aloha: simpler, no synchronization
- when frame first arrives
  - transmit immediately
- collision probability increases:
  - frame sent at $t_0$ collides with other frames sent in $[t_0-1, t_0+1]$
Pure Aloha efficiency

\[ P(\text{success by given node}) = P(\text{node transmits}) \cdot \]
\[ P(\text{no other node transmits in } [p_0-1, p_0]) \cdot \]
\[ P(\text{no other node transmits in } [p_0-1, p_0]) \]
\[ = p \cdot (1-p)^{N-1} \cdot (1-p)^{N-1} \]
\[ = p \cdot (1-p)^{2(N-1)} \]

... choosing optimum p and then letting n → infty ...

Even worse!

\[ = 1/(2e) = .18 \]
CSMA (Carrier Sense Multiple Access)

**CSMA**: listen before transmit:
- If channel sensed idle: transmit entire frame
- If channel sensed busy, defer transmission

- Human analogy: don’t interrupt others!
CSMA collisions

Collisions can still occur:
Propagation delay means two nodes may not hear each other’s transmission.

Collision:
Entire packet transmission time wasted.

Note:
Role of distance & propagation delay in determining collision probability.
CSMA/CD (Collision Detection)

CSMA/CD: carrier sensing, deferral as in CSMA
  - collisions detected within short time
  - colliding transmissions aborted, reducing channel wastage

Collision detection:
  - easy in wired LANs: measure signal strengths, compare transmitted, received signals
  - difficult in wireless LANs: receiver shut off while transmitting

Human analogy: the polite conversationalist
CSMA/CD collision detection
“Taking Turns” MAC protocols

channel partitioning MAC protocols:
- share channel efficiently and fairly at high load
- inefficient at low load: delay in channel access, 1/N bandwidth allocated even if only 1 active node!

Random access MAC protocols
- efficient at low load: single node can fully utilize channel
- high load: collision overhead

“taking turns” protocols
- look for best of both worlds!
“Taking Turns” MAC protocols

Polling:
- master node “invites” slave nodes to transmit in turn
- concerns:
  - polling overhead
  - latency
  - single point of failure (master)

Token passing:
- control token passed from one node to next sequentially.
- token message
- concerns:
  - token overhead
  - latency
  - single point of failure (token)
Summary of MAC protocols

What do you do with a shared media?

- Channel Partitioning, by time, frequency or code
  - Time Division, Code Division, Frequency Division
- Random partitioning (dynamic),
  - ALOHA, S-ALOHA, CSMA, CSMA/CD
  - carrier sensing: easy in some technologies (wire), hard in others (wireless)
  - CSMA/CD used in Ethernet
- Taking Turns
  - polling from a central site, token passing
LAN technologies

Data link layer so far:
- services, error detection/correction, multiple access

Next: LAN technologies
- addressing
- Ethernet
- hubs, bridges, switches
- 802.11
- PPP
- ATM
LAN Addresses and ARP

32-bit IP address:
- network-layer address
- used to get datagram to destination IP network (recall IP network definition)

LAN (or MAC or physical or Ethernet) address:
- used to get datagram from one interface to another physically-connected interface (same network)
- 48 bit MAC address (for most LANs) burned in the adapter ROM
LAN Addresses and ARP

Each adapter on LAN has unique LAN address

![Diagram showing LAN addresses and adapters](image)
LAN Address (more)

- MAC address allocation administered by IEEE
- Manufacturer buys portion of MAC address space (to assure uniqueness)
- Analogy:
  - (a) MAC address: like Social Security Number
  - (b) IP address: like postal address
- MAC flat address ⇒ portability
  - can move LAN card from one LAN to another
- IP hierarchical address NOT portable
  - depends on IP network to which node is attached
Recall earlier routing discussion

Starting at A, given IP datagram addressed to B:
- look up net. address of B, find B on same net. as A
- link layer send datagram to B inside link-layer frame
ARP: Address Resolution Protocol

Question: how to determine MAC address of B knowing B's IP address?

- Each IP node (Host, Router) on LAN has ARP table.
- ARP Table: IP/MAC address mappings for some LAN nodes.
  - IP address; MAC address; TTL

  - TTL (Time To Live): time after which address mapping will be forgotten (typically 20 min).
ARP protocol

- A wants to send datagram to B, and A knows B's IP address.
- Suppose B's MAC address is not in A's ARP table.
- A broadcasts ARP query packet, containing B's IP address
  - all machines on LAN receive ARP query
- B receives ARP packet, replies to A with its (B's) MAC address
  - frame sent to A's MAC address (unicast)
- A caches (saves) IP-to-MAC address pair in its ARP table until information becomes old (times out)
  - soft state: information that times out (goes away) unless refreshed
- ARP is "plug-and-play":
  - nodes create their ARP tables without intervention from net administrator
Routing to another LAN

walkthrough: send datagram from A to B via R

assume A knows B IP address

- Two ARP tables in router R, one for each IP network (LAN)

![Diagram showing routing from A to B via R]
- A creates datagram with source A, destination B
- A uses ARP to get R’s MAC address for 111.111.111.110
- A creates link-layer frame with R’s MAC address as dest, frame contains A-to-B IP datagram
- A’s data link layer sends frame
- R’s data link layer receives frame
- R removes IP datagram from Ethernet frame, sees its destined to B
- R uses ARP to get B’s physical layer address
- R creates frame containing A-to-B IP datagram sends to B
Ethernet

“dominant” LAN technology:
- cheap $20 for 100Mbs!
- first widely used LAN technology
- Simpler, cheaper than token LANs and ATM
- Kept up with speed race: 10, 100, 1000 Mbps

Metcalfe’s Ethernet sketch
Ethernet Frame Structure

Sending adapter encapsulates IP datagram (or other network layer protocol packet) in Ethernet frame

Preamble:
- 7 bytes with pattern 10101010 followed by one byte with pattern 10101011
- used to synchronize receiver, sender clock rates
Ethernet Frame Structure (more)

- **Addresses**: 6 bytes
  - if adapter receives frame with matching destination address, or with broadcast address (e.g., ARP packet), it passes data in frame to net-layer protocol
  - otherwise, adapter discards frame

- **Type**: indicates the higher layer protocol, mostly IP but others may be supported such as Novell IPX and AppleTalk

- **CRC**: checked at receiver. If error is detected, the
Unreliable, connectionless service

- **Connectionless**: No handshaking between sending and receiving adapter.
- **Unreliable**: receiving adapter doesn’t send acks or nacks to sending adapter
  - stream of datagrams passed to network layer can have gaps
  - gaps will be filled if app is using TCP
  - otherwise, app will see the gaps
Ethernet uses CSMA/CD

- No slots
- adapter doesn’t transmit if it senses that some other adapter is transmitting, that is, carrier sense
- transmitting adapter aborts when it senses that another adapter is transmitting, that is, collision detection

- Before attempting a retransmission, adapter waits a random time, that is, random access
Ethernet CSMA/CD algorithm

1. Adaptor gets datagram from and creates frame
2. If adapter senses channel idle, it starts to transmit frame. If it senses channel busy, waits until channel idle and then transmits
3. If adapter transmits entire frame without detecting another transmission, the adapter is done with frame!
4. If adapter detects another transmission while transmitting, aborts and sends jam signal
5. After aborting, adapter enters exponential backoff: after the mth collision, adapter chooses a K at random from \{0,1,2,\ldots,2^m-1\}. Adapter waits K*512 bit times and returns to Step 2
Ethernet's CSMA/CD (more)

Jam Signal: make sure all other transmitters are aware of collision; 48 bits;

Bit time: .1 microsec for 10 Mbps Ethernet;
for K=1023, wait time is about 50 msec

Exponential Backoff:
- **Goal**: adapt retransmission attempts to estimated current load
  - heavy load: random wait will be longer
- first collision: choose K from \{0,1\}; delay is K \times 512 bit transmission times
- after second collision: choose K from \{0,1,2,3\}...
- after ten collisions, choose K from \{0,1,2,3,4,\ldots,1023\}

See/interact with Java applet on AWL Web site: highly recommended!
CSMA/CD efficiency

- $T_{prop} = \text{max prop between 2 nodes in LAN}$
- $t_{trans} = \text{time to transmit max-size frame}$

\[
\text{efficiency} = \frac{1}{1 + 5\frac{T_{prop}}{t_{trans}}}
\]

- Efficiency goes to 1 as $T_{prop}$ goes to 0
- Goes to 1 as $t_{trans}$ goes to infinity
- Much better than ALOHA, but still decentralized, simple, and cheap
Ethernet Technologies: 10Base2

- **10**: 10Mbps; **2**: under 200 meters max cable length
- thin coaxial cable in a bus topology
  - transmits packet travels in both directions

- repeaters used to connect up to multiple segments
- repeater repeats bits it hears on one interface to its other interfaces: physical layer device only!
- has become a legacy technology
10BaseT and 100BaseT

- 10/100 Mbps rate; latter called “fast ethernet”
- T stands for Twisted Pair
- Nodes connect to a hub: “star topology”; 100 m max distance between nodes and hub

- Hubs are essentially physical-layer repeaters:
  - bits coming in one link go out all other links
  - no frame buffering
  - no CSMA/CD at hub: adapters detect collisions
  - provides net management functionality
Manchester encoding

- Used in 10BaseT, 10Base2
- Each bit has a transition
- Allows clocks in sending and receiving nodes to synchronize to each other
  - no need for a centralized, global clock among nodes!
- Hey, this is physical-layer stuff!
**Gbit Ethernet**

- use standard Ethernet frame format
- allows for point-to-point links and shared broadcast channels
- in shared mode, CSMA/CD is used; short distances between nodes to be efficient
- uses hubs, called here “Buffered Distributors”
- Full-Duplex at 1 Gbps for point-to-point links
- 10 Gbps now!
Chapter 5 outline

- 5.1 Introduction and services
- 5.2 Error detection and correction
- 5.3 Multiple access protocols
- 5.4 LAN addresses and ARP
- 5.5 Ethernet
- 5.6 Hubs, bridges, and switches
- 5.7 Wireless links and LANs
- 5.8 PPP
- 5.9 ATM
- 5.10 Frame Relay
Interconnecting LAN segments

- Hubs
- Bridges
- Switches
  - Remark: switches are essentially multi-port bridges.
  - What we say about bridges also holds for switches!
Interconnecting with hubs

- Backbone hub interconnects LAN segments
- Extends max distance between nodes
- But individual segment collision domains become one large collision domain
  - if a node in CS and a node EE transmit at same time: collision
- Can't interconnect 10BaseT & 100BaseT
Bridges

- Link layer device
  - stores and forwards Ethernet frames
  - examines frame header and selectively forwards frame based on MAC dest address
  - when frame is to be forwarded on segment, uses CSMA/CD to access segment
- transparent
  - hosts are unaware of presence of bridges
- plug-and-play, self-learning
  - bridges do not need to be configured
Bridges: traffic isolation

- Bridge installation breaks LAN into LAN segments
- Bridges filter packets:
  - same-LAN-segment frames not usually forwarded onto other LAN segments
  - segments become separate collision domains

![Diagram of bridges and LAN segments]

- collision domain
- bridge
- collision domain
- LAN segment
- LAN (IP network)
- = hub
- = host
Forwarding

How do determine to which LAN segment to forward frame?
• Looks like a routing problem...
Self learning

- A bridge has a bridge table
- entry in bridge table:
  - (Node LAN Address, Bridge Interface, Time Stamp)
  - stale entries in table dropped (TTL can be 60 min)
- bridges learn which hosts can be reached through which interfaces
  - when frame received, bridge “learns” location of sender: incoming LAN segment
  - records sender/location pair in bridge table
Filtering/Forwarding

When bridge receives a frame:

index bridge table using MAC dest address
if entry found for destination
    then{
        if dest on segment from which frame arrived
            then drop the frame
        else forward the frame on interface indicated
    }
else flood
    forward on all but the interface on which the frame arrived
Bridge example

Suppose C sends frame to D and D replies back with frame to C.

- Bridge receives frame from from C
  - notes in bridge table that C is on interface 1
  - because D is not in table, bridge sends frame into interfaces 2 and 3

- frame received by D
Bridge Learning: example

- D generates frame for C, sends
- bridge receives frame
  - notes in bridge table that D is on interface 2
  - bridge knows C is on interface 1, so selectively forwards frame to interface 1