LAN devices

Fulvio Risso
Politecnico di Torino
LAN devices

- **L1**: Repeater
  - Hub
  - Separate physical domains

- **L2**: Bridge
  - Switch
  - Separate collision domains

- **L3**: Router
  - L3 switch
  - Separate broadcast domains
  - Not really specific for LANs
  - Not covered in the current slides
Repeater (1)

- Receives and propagates a sequence of bits
- Interconnection at the physical layer
  - E.g. fiber to copper
  - Same MAC (i.e., all ports must have the same speed)
  - Recovers signal degradation (long cables), allowing larger distances

![Diagram of network layers]
Repeater (2)

- Functions
  - Signal Amplification
  - Signal Symmetry
  - Signal Retiming
  - Carrier Sense and Data Repeat
  - Collision Detection and Jam Generation
  - Test functions

- Active device
  - Not just a “passive” backplane (i.e., signal amplification)

- No longer used (at least on Ethernet LANs)
Repeaters and Hubs

- Classical repeater: 2 port devices
- Multiport repeater: repeater with more than 2 ports
  - Known as “hub”
  - It became common with the adoption of structured cabling
  - More flexible (and robust) than the old coax cable
- Extends the collision domain of Ethernet
  - Max diameter unchanged
  - More details in the next slides
Bridge (1)

- Introduced by DEC in 1983 (LANBridge 100)
  - Pure software
  - 2 ports (mainly for economic reasons)
- Interconnection at the data-link layer
  - E.g. Ethernet to WiFi, Ethernet to Fast Ethernet
  - Different MACs (medium access mechanism, framing)
Bridge (2)

- Original objective
  - Interconnection between different LANs
    - In theory; in practice it is often impossible due to MTU issues (data-link does not have fragmentation)
  - LAN extension (total diameter)
    - Especially useful for FastEthernet and upper speed (200m)
- Works by receiving and re-transmitting (later) a frame
  - Usually stores the frame
  - It may modify the frame sent out (e.g. Ethernet to Token Ring)

---

**Ethernet DIX**

<table>
<thead>
<tr>
<th>Preamble</th>
<th>SFD</th>
<th>MAC Dest.</th>
<th>MAC Source</th>
<th>Length</th>
<th>Data</th>
<th>FCS</th>
<th>IFG</th>
</tr>
</thead>
</table>

---

**802.5 Token Ring**

<table>
<thead>
<tr>
<th>Starting Delimiter</th>
<th>Access Control</th>
<th>Frame Control</th>
<th>MAC Dest.</th>
<th>MAC Source</th>
<th>Routing Inform.</th>
<th>Data</th>
<th>FCS</th>
<th>Ending Delimiter</th>
<th>Frame Status</th>
</tr>
</thead>
</table>
Collision and broadcast domains (1)

Collision Domain 1

Collision Domain 2

Collision Domain 3

Collision Domain 4

Broadcast domain

"Physical" LAN

"Logical" LAN
Collision and broadcast domains (2)

- Collision domain: area where a single instance of CSMA/CD operates
  - I.e., the area covered by a single “physical” link
  - Frames are immediately propagated over all the links (possibly through repeaters)

- Broadcast domain: area where frames can be propagated
  - I.e., the area on which a LAN operates
  - Can include several collision domains
  - Frames can be stored and later propagated over the other collision domains
Collision and broadcast domains (3)

- Repeaters extend the collision domain
- Bridges create different collision domains and extend the broadcast domain
  - I.e., bridges decouple broadcast domain from collision domain
- Please note that a LAN is a *shared communication medium*
  - But... not necessarily at the physical layer
  - This feature can be emulated at logical level

- The splitting between collision and broadcast domains open the door to NICs operating in “full duplex” mode
Full duplex mode

- Possible operating mode of network interfaces
  - Half Duplex: RX and TX cannot happen at the same time
  - Full Duplex: allows RX and TX cables to be used at the same time
- The collision domain is limited to each (two) directly connected station
- Available when the other party can *temporarily store* the frame
  - E.g., instead of just repeating the bytes on the other ports, such as a repeater does
  - Examples: host ↔ host, host ↔ bridge, bridge ↔ bridge
- Introduced with Fast Ethernet (part of 802.3x)
Full duplex mode: advantages

- **Bandwidth**
  - In theory, throughput x2
  - In practice, minimum advantage for clients and servers
    - Clients tend to saturate downlinks, servers uplinks
  - May be interesting for bridges on the backbone
    - More symmetrical bandwidth

- **CSMA/CD**
  - No longer needed, since collisions are no longer possible
    - With CSMA/CD, TX and RX together are used to detect collisions
  - Advantages
    - No requirement for min frame size for Ethernet
    - No limits on the network size on Ethernet (no collision domain)
Bridge: smart forwarding process

- Bridges, in principle, should propagate each frame in the entire network
  - This seems to be a requirement of the "shared communication mode"
  - But... is this useful?
  - If a frame is directed to Host A, why should I send it also to Host B?

- By-product: if the bridge has a smart forwarding process, it can implement traffic segregation
  - Increase the aggregate bandwidth of the network
  - Right now, the most important reason for using these devices
  - Forwarding technique based on MAC Destination address
Smart forwarding process (1)

1. A → E
2. A → E
3. E → F
4. C → D

Diagram showing the network with nodes A, B, C, D, E, F and the connections between them.
Smart forwarding process (2)

- A bridge can implement the “store and forward” technology
  - Receives a frame on one interface
  - Stores the frame into a local buffer
  - Analyzes the destination address
    - Why the MAC Destination precedes the MAC Source in Ethernet frames?
  - Forward it on the right port (if needed)

- Ports belong to different collision domain, therefore a bridge can send/receive traffic at the same time over different ports
  - Bridges have buffers in order to absorb bursts and to wait for the proper transmission slot
Smart forwarding process (3)

- Smarter forwarding rules
  - Unicast: only on the port toward we can reach the destination (Destination MAC-based forwarding)
  - Multicast, Broadcast: flooding
    - All ports except the port on which the frame has been received

- Transient
  - If the MAC address is not present in the MAC forwarding table
  - Bridge looks like an hub
    - Frame duplicated on all ports except the one on which it was received
    - Different from hubs because of the delayed forwarding (store and forward algorithm), while hubs forward \textit{bits} immediately

- A MAC forwarding table must be available locally
Forwarding process (unicast and broadcast)

1. **Begin**
2. **Received frame on port X**
   - **Errors (collision, CRC)?**
     - **Y**
       - **MAC destination in DB?**
         - **Y**
           - **Destination port == X?**
             - **Y**
               - **Discard frame**
             - **N**
               - **Forward on selected port**
     - **N**
       - **Forward on all ports (except X)**
   - **N**
     - **End**
Transparent bridge

- IEEE standardized the bridging function in 802.1D
- 802.1D defines Transparent Bridges
  - Other (non transparent) bridges have been proposed in the past (e.g. Token Ring networks)
    - No longer in use
  - Transparent bridges have been proposed in Ethernet
- Transparency
  - End systems must operate in the same way (same frames, some format, etc) with or without bridges
  - Performance (throughput, max distances) may be different from the original network, but functionalities are the same
Changes in sent/received frames on hosts (1)

- Changes in sent frames
  - Nothing at all

- Changes in received frames at the NIC level
  - No receiving all frames anymore
  - All the frames sent/received on the current network segment, all broadcast/multicast, plus the ones with its MAC address
    - At limit (host with point-to-point connection to a bridge): all broadcast/multicast, plus the ones with its MAC address
    - Transient excluded

- May become useless
  - MAC filtering on the NIC
  - NIC in promiscuous mode
Changes in sent/received frames on hosts (2)

- Changes in received frames at the OS level
  - Nothing at all
  - Bridges filters frames that were previously filtered by the NIC
    - The result at the OS level is the same
Transparent bridges and port addresses

- The port of a bridge has a MAC level and therefore it may have a MAC address
  - Often bridges have an unique ID for the entire device
- That MAC address is never used when forwarding data frames
- It is used when frames are generated/received by the switch itself
  - E.g. management frames
Transparent bridges: new components

- In order to operate successfully, a transparent bridge requires three additional components:
  - A local forwarding table (*filtering database*)
  - Stations auto-learning (*backward learning*)
  - Loop detection (*spanning tree algorithm*)
The Filtering Database (1)

- Table with the “location” of any MAC address found in the network
  - MAC address
  - Destination port
  - Ageing time (default expire after 300 s)
  - Port status (depending by spanning tree protocol)
- Required for the “smart forwarding process”
The Filtering Database (2)

- Entry types
  - Dynamic
    - Populated and updated by the backward learning process
    - Max entries: $2 \div 64$ K
  - Static
    - Not updated by the learning process
    - Usually < 1K entries
- Old dynamic entries are purged out of the filtering database
  - E.g., stations that do no longer exist on the network
  - Default: 300 seconds
### Filtering database: example

```bash
cisco-switch-1> show cam dynamic
```

* = Static Entry. + = Permanent Entry.
# = System Entry X = Port Security Entry

<table>
<thead>
<tr>
<th>Dest MAC Address</th>
<th>Ports</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>00-00-86-1a-a6-44</td>
<td>1/1</td>
<td>1</td>
</tr>
<tr>
<td>00-00-c9-10-b3-0f</td>
<td>1/1</td>
<td>0</td>
</tr>
<tr>
<td>00-00-f8-31-1c-3b</td>
<td>1/2</td>
<td>4</td>
</tr>
<tr>
<td>00-00-f8-31-f7-a0</td>
<td>1/1</td>
<td>2</td>
</tr>
<tr>
<td>00-01-e7-00-e3-80</td>
<td>2/2</td>
<td>0</td>
</tr>
<tr>
<td>00-02-a5-84-a7-a6</td>
<td>2/1</td>
<td>1</td>
</tr>
<tr>
<td>00-02-b3-1e-b4-aa</td>
<td>2/1</td>
<td>5</td>
</tr>
<tr>
<td>00-02-b3-1e-da-da</td>
<td>2/5</td>
<td>1</td>
</tr>
<tr>
<td>00-02-b3-1e-dc-fd</td>
<td>2/4</td>
<td>2</td>
</tr>
</tbody>
</table>
Backward learning (1)

MAC Filtering Database

<table>
<thead>
<tr>
<th>MAC</th>
<th>Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>2</td>
</tr>
</tbody>
</table>

1. A → E
2. A → E
3. C → D
4. E → F
Backward learning (2)

- Bridges learn the topology by inspecting received frames
  - Analysis of MAC source address
- Works also in presence of multiple bridges
  - Remote bridges learn the position anyway, even if the end-system is not connected locally
Backward learning (3)

Begin

Received frame on port X

Source MAC address found in the DB? N Y

Update port and ageing time

Add new entry in DB

End

Background process: Discard zombies
Moving end systems (1)

- If the end-system generates broadcast frame immediately
  - No problems

- Diagram showing network traffic and filtering database entries with MAC addresses and ports.
Moving end systems (2)

- If the end-system generates unicast traffic immediately
  - We may have forwarding errors
    - D → B is correctly delivered
    - C → B is lost
Moving end systems (3)

- If the end-system does not generate traffic at all
  - We may have forwarding troubles
    - D → B is correctly delivered
      - The frame is forwarded also to the original destination
  - C → B is lost

![Diagram](image_url)
Moving end-systems (4)

- Broadcast frame
  - Receives all the network, therefore all the bridges update the location of the current station

- Unicast frame
  - Potentially reaches only a portion of the network, hence the rest may still have the old location of the station

- In the real world
  - Windows host typically generates a lot of broadcast
    - No problems when moving from one place to another
  - UNIX servers and virtualized hosts (e.g., Vmware) are often silent if not solicited
    - Need to wait for the aging time
Moving end systems (5)

- The aging time
  - Usually enough in order to cope with manual movements
    - A laptop moved from office to lab
  - Represents the worst-case black-out time for an end system

- Some problems may appear in specific environments
  - E.g. fault-tolerant NICs
    - We need to react much quickly than 5min
  - NIC driver has to generate an additional broadcast frame
L2 networks and mobility

- L2 networks natively support mobility
  - Just enough to send some broadcast immediately
- Mobility at L3 level requires MobileIP, which is often not available on real devices

Lesson learned

- Much easier to move an host at L2 level and keep all the connection active (the IP address does not change)...
- ... than move an host at L3 level (which requires to change the IP address)
Some notes about filtering database

- Please note that...
  - An end-system whose MAC address is not in the DB is *always reachable*
    - Corollary: a frame sent to a non-existing host will always be forwarded in all the network
  - An end-system whose MAC address is in the DB may be *unreachable*
Possible attacks to the filtering database

- MAC Flooding Attack
  - Generation of frames with random MAC sources
  - Filtering database gets full
  - Bridges will start flooding most of the frames
    - All the ones whose destination address is not present in the DB
- Objectives
  - Forces bridges to operate like hubs, so that we can intercept traffic generated by other stations
    - Slows down the network
  - Some vendors give the opportunity to limit the number of MAC address learnt on each port
Possible attacks to the filtering database (2)

- Packet storms
  - Generation of frames to non-existing stations
  - Frames are always send to the entire network
- Objective
  - Slows down the network
Bridges and meshes

- Two problems
  - Frames can enter in a loop
  - Backward learning no longer able to operate

- It’s now the time to present the third component (i.e. “Spanning Tree”) after the ones we presented earlier
  - “Filtering Database” and “Backward Learning”
Bridges and meshes: the loop problem (1)
Bridges and meshes: the loop problem (2)

- Frames can enter in a loop
  - Multicast/broadcast frames
    - Very common
  - Frame to a non-existing station
    - MAC address not present in the filtering DB (e.g. non existing station)
    - Problem that may happen rarely (unless under attack)
      - IP sends an ARP before contacting an L2 station
      - If the station does not exist, the ARP will never get a reply and the destination MAC address is unknown
      - Therefore, no MAC frames will be sent to that station intentionally
The Broadcast Storm

- Massive load due to broadcast/multicast traffic on a LAN
- One of the most dangerous problems at data-link layer
- No solutions, except for disabling (physically) loops
  - E.g., detach a cable from a bridge
- Network operators are almost impotent in such this case
- Due to the lack of a “time-to-live” field in L2 frames
- L3 networks can tolerate transient loops
  - TTL available on L3 packets
Bridges and meshes: the learning problem (1)
Bridges and meshes: the learning problem (2)

- Backward learning problem
  - Switches may have inconsistent filtering database
  - An entry in the filtering database may change the port indefinitely
    - An entry may not able to reach a stable state
    - Transient loops can be created among back-to-back bridges
      - B1 forwards to B2 that forwards to B1m,...
      - Larger (B1-B2-B3-B1) loops may occur as well
The Spanning Tree idea: no loops in the network
Spanning Tree

- In order to avoid troubles, you must avoid loops in the physical network
  - Either create loop-free networks
    - Discouraged; not robust
  - Or define an algorithm that disables (temporarily) loops

- 802.1D
  - Original idea from Radia Perlman, PhD @DEC

- Meshes detected and disabled; the network becomes a tree
  - Unique path between any source and any destination

- Operates periodically (every second)
  - Decides which port set to forwarding state and which port set to blocking state

- More details in another set of slides
Bridges

Advantages

- Transparent to the network stack (and to the application)
- Works with all L3 protocols
- Automatic configuration (although not optimal)
- Allows automatic reconfiguration in case of faults
- Increases performance (different collision domains)

Problems

- Not suitable for complex networks (e.g. WAN)
- No filtering for broadcast frames
- No load balancing over multiple parallel links
Bridge architecture

- Backward Learning
- Spanning Tree Protocol
- Filtering database
- Forwarding process

A → B
Bridges and switches (1)

- Bridge
  - Originally 2 ports, then more
  - Software-based architecture
  - No longer used in real networks
  - Still some PC-based implementations
    - For research or some special purpose
  - WiFi access points are bridges
Bridges and switches (2)

- Switch
  - Same device, different technology
  - Hardware based forwarding and learning
  - Lookup through CAMs (Content Addressable Memories)
  - Spanning Tree in software
    - Convergence time in several seconds, hence hardware implementation is useless
  - Can implement a “cut-through” forwarding technology
    - A frame can be forwarded on the target port immediately after receiving the Destination MAC
      - The destination port must be free at that time
    - Faster than “store and forward”
    - Requires all ports operating at the same speed
Switch internals

- Shared bus or switching matrix
  + speed
  - complexity

- No CSMA/CD (Full Duplex)
  + speed
  - useful only on some links (e.g., intra-switch)

- Queuing system (often on the output link)
  + decoupling of different physical speed
  + absorbs burst
  - can drop frames

- Central CPU and memory
  + intelligence
  - configuration
  - bugs

- Filtering Database
  + efficient lookup
  - Table may become full
  - transient

<table>
<thead>
<tr>
<th>Port</th>
<th>Host</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AA-BB...</td>
</tr>
<tr>
<td>1</td>
<td>AA-BC...</td>
</tr>
<tr>
<td>2</td>
<td>AA-BD...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Switched LANs (1)

- Progressive replacement of shared segments with switches
Switched LANs (2)

Data center (CED)

Internet
Switched LANs (3)

- Currently, end systems directly connected to switches
  - More aggregated bandwidth
  - No need to replace NIC on clients when moving from hubs to switches
  - Switches may be 10/100/1000 and support different speed on the client side
    - Possibility to smooth upgrade of the network (NICs, hubs/switches), mixing different Ethernet technologies
    - Hub did not support multiple speed
Switched LANs (4)

- Technologies no longer used in the real world of Ethernet
  - CSMA/CD
    - Only one station can be attached to a physical link (no need to arbitrate the channel)
  - Frame bursting
  - Carrier Extension

- What remains in the Ethernet
  - Framing

- Maximum diameter of an Ethernet (Fast/Giga/...) network
  - Max diameter (for collision domain) is no longer a problem
  - Max cable length (due to signal attenuation) is still a problem
    - E.g., 100m from end-system to a switch (twisted pair) is still a valid limit
Switched networks and throughput (1)

- Aggregate bandwidth increases

/!
Throughput may not !
/

1) Uplink speed is a critical factor
   - Uplinks must sustain the traffic of all the attached station
   - Links toward servers must be fast enough
   - Is it a good choice to have clients connected at 1Gbps?

2) Poor segregation of different network segments
   - Backward learning process generates transient
   - Possibility to attack the network (flooding attacks, ...)


Switched networks and throughput (2)

3) Buffers play an important role in switches
   - Classical Ethernet implements a “reliable” transmission
     - Why can CIFS and NFS use UDP for data transfers?
   - Switches may drop frames due to congestions (limited buffer size)
   - TCP algorithms (timeout, fast retransmit, ...) come into play
     - Dramatic decline in throughput
     - TCP dimensioned for reacting in about hundred ms, not microseconds
Switched networks and scalability

- Broadcast (and, in some sense, multicast) is still a problem
  - Single broadcast domain
- Network size limited by the number of stations
  - Usually, no more than 1000 stations
Switched networks and security

- Poor segregation of different network segments
  - Backward learning process generates transient
  - Possibility to attack the network (flooding attacks, ...)
  - No “hard” way to segregate traffic in multiple segments
    - Unless VLANs are used
Conclusions

- Repeaters, Hubs and Bridges are historical
- Switches are really used
  - Most real network are now “pure switched networks”
  - Increased performance
  - Departure from some peculiar characteristics of a LAN
    - E.g., low error rate
    - No CSMA/CD, large diameters (in Ethernet)
- Components of a switch
  - Filtering Database
  - Learning process
  - Spanning Tree