Data Link Layer

The data link layer uses physical addresses assigned to each physical network device in the local network to route data from one physical device to another. These addresses are called Media Access Control (MAC) addresses in TCP/IP.

The data link layer has the following features:

✦ Receives each packet from the network layer on the sending host

✦ Wraps up the packet in a data frame along with local routing data(the physical MAC address)

✦ Sends the data frame to the physical layer to code an electrical or optical signal

The physical layer transmits the data frame over a wire or over the air

(wireless transmission).

On the receiving host, the data link layer does the following:

✦ Unwraps the data frame received

✦ Extracts the packet out of the data frame

✦ Sends the packet up to the network layer

Hence, whenever you consider Layer 2 in TCP/IP, you really need to think about Ethernet and physical MAC addresses. Layer 2, the data link layer, is only concerned with local-area networks (LANs)

Switches

Layer 2 switch is a network device that creates one collision domain per port and forwards data frames only on the outbound port that reaches the destination of the frame.

Switch characteristics are as follows:

✦ Switches are faster than bridges because they use hardware ASICs instead of software to perform their operations.

✦ Switches are typically faster than routers because they do not need to look at the network layer (Layer 3) IP packet header. They only inspect the data-link (Layer 2) frame to look at the source and destination MAC address of the frame. This is why they are called Layer 2 switches: They only operate on the data-link (Layer 2) frame.

Hubs, Bridges, and Switches			
Feature	*Hubs*	*Bridges*	*Switches*
Technology	Port multiplexing	Software switching	ASIC switching
Duplex	Half	Half	Half/Full
Speed	Turtle	Bear	Leopard
VLAN support	No	No	Yes
Collision domain	Whole hub	1 per port	1 per VLAN
Broadcast domain	Whole hub	Whole bridge	1 per VLAN

Basic Switch Functions

A Layer 2 switch must accomplish three tasks:

✦ Learn about the MAC addresses of devices connected to the switch

✦ Decide whether to forward frames it receives from host devices or other switches

✦ Avoid creating any Layer 2 loops

Address learning

Layer 2 switches learn the MAC addresses of devices connected to the switch as follows:

The switch inspects each data frame that enters the switch. It saves the port number where the frame entered along with the source MAC address of that frame. The MAC address and the corresponding port number are saved in a MAC address table.

As devices send frames into the switch, the switch slowly builds a complete MAC address table that contains

• The MAC address of each device connected

• The port number through which that device is sending frames into the switch

Flooding, forwarding, and filtering frames

Layer 2 switches need to decide whether to forward frames they receive. They also need to figure out over which outbound port they forward the frame. To do this, switches use the MAC address table:

1. The switch inspects each frame that enters on an inbound port: It looks at the source and destination MAC address of the frame.

2. The switch searches for the destination MAC address of the frame in the MAC address table:

• If the switch finds the destination, the switch forwards the frame on the outbound port saved in the MAC address table for that destination MAC address.

• If the switch does not find the destination MAC address in its MAC address table, it forwards the frame out on all outbound ports except on the one through which that frame come in.

Terms and their definitions

Address learning: The process by which a Layer 2 switch learns which MAC address is connected to each switch port. The process involves the switch saving the port number where each frame enters along with the source MAC address of that frame. The incoming port number and the MAC address are saved in a MAC address table.

Flooding: The process by which a Layer 2 switch forwards a frame out on all outbound ports except on the port through which that frame came in. A switch floods a frame whenever it does not find the frame's destination MAC address in the MAC address table. It also floods broadcast frames.

Forwarding: The process by which a Layer 2 switch forwards a frame on the outbound port saved in the MAC address table for that destination MAC address.

Filtering: The process by which a Layer 2 switch discards a frame without sending it out on any ports. A switch filters a frame whenever the frame's source MAC address and frame's destination MAC address are registered to the same switch port, according to the MAC address table.

Avoiding loops: The process by which Layer 2 switches eliminate trans- mission loops created by redundant interswitch links. The Spanning Tree Protocol (STP) handles this process.

Broadcast storm: A broadcast frame that bounces forever among switches interconnected with redundant links. Broadcast storms waste bandwidth and may thrash the MAC address table.

MAC address table thrashing: Multiple ports attached to the same MAC address. The switch does not know on which outgoing port it can reach that MAC address.

Unicast transmission: Involves a device sending a frame to a single target device.

Multicast transmission: Involves a device sending a frame to multiple target devices.

Broadcast transmission: Involves a device sending a frame to all devices in its local network.

SPANNING TREE PROTOCOL (STP)

Spanning Tree Protocol (STP) that is used by Layer 2 switches to avoid loops in topologies with redundant Interswitch links.

✦ Broadcast storms: A broadcast frame that bounces forever between switches interconnected with redundant links is called a broadcast storm. Broadcast storms waste bandwidth and may thrash the MAC address table.

✦ MAC address table thrashing: The MAC address table gets thrashed when multiple ports are attached to the same MAC address. The switch does not know on which outgoing port it can reach that MAC address.

So, how exactly does STP eliminate loops in LANs?

1. STP basically monitors the network and catalogs each link, particularly redundant links.
2. Next, STP disables redundant links, setting up preferred, optimized links between switches. The preferred, optimized links are used under normal circumstances. Should any of the preferred links fail, one of the nonpreferred redundant links is enabled and used instead.
3. STP assigns a root bridge that acts sort of like the decision maker in the network. The term root bridge typically refers to a Layer 2 switch. It can also really be a bridge, but more commonly it is a switch. The root bridge decides which routes are preferred and which routes are nonpreferred.
4. The root bridge interacts with nonroot bridges: other switches in the LAN. STP is enabled on all switches in the LAN. The root bridge and the nonroot bridges have specific roles within STP.
5. Switch ports are categorized by whether they forward frames and whether they are the endpoints of a preferred, optimized link within the LAN.

STP Operation Flow

The Spanning Tree Protocol executes three operations to achieve a stable, loop-free LAN:

+ Electing STP root bridge: The switch with the lowest bridge ID is elected the root bridge.

+ Assigning STP port types: Each port on every switch in the LAN is assigned a type that defines its behavior. Switches communicate using STP to assign a type to their ports. This determines the ports' behavior to be either forwarding (that is, the port forwards data-link frames) or blocking (that is, the port sends no data-link frames).

+ Achieving STP convergence: Convergence is the result of assigning port types. After switches have assigned a forwarding type or a blocking type to all ports that interconnect them, STP achieves a stable, loop-free LAN. I now look at each of these steps in detail.

In detail:

Electing a root bridge

STP's first step is to choose a root bridge. STP relies on the bridge ID to decide which switch should become the root bridge. STP chooses the switch with the lowest bridge ID to be the root bridge.

Bridge 1D

The bridge ID is a number composed of the switch's MAC address and its STP priority. The STP priority on any given Layer 2 switch is by default 32768. So, because the priority of all switches is by default the same (32768), STP really chooses the switch with the lowest MAC address to be the root bridge. The switch with the lowest MAC address is chosen as the root bridge.

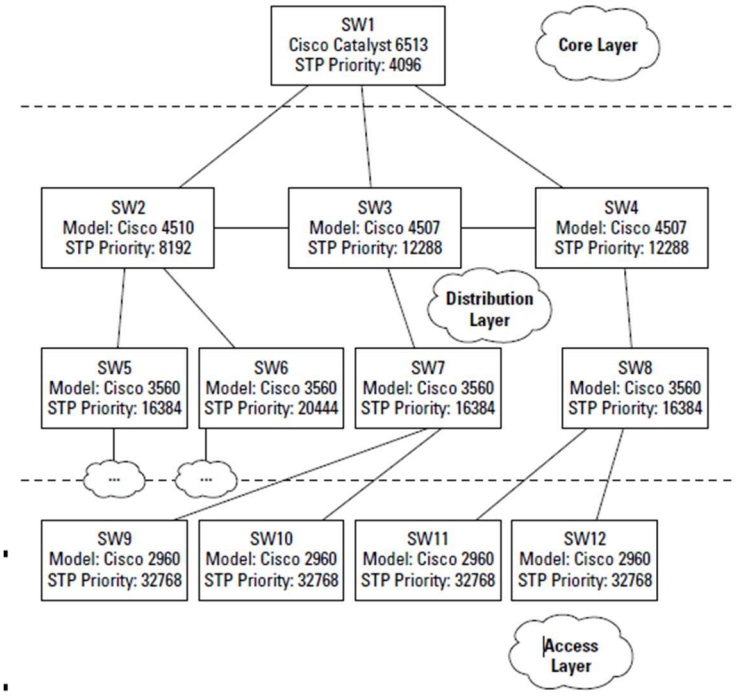
STP priorityų

What if the switch with the lowest MAC address is actually an old, slow switch? You wouldn't want the slowest switch in your LAN to become the root bridge because it will slow every other switch in your network. You can use the STP priority setting to control which switch will have the lowest bridge ID.

Best practice is to set the STP priority to a low value (smaller than the default 32768 STP priority) on the fastest switch in your LAN. This causes STP to choose the fastest switch in your LAN as the root bridge.

You can set the STP priority to any number between 0 and 61440. However, the STP priority value has to be a multiple of 4096. So, the STP priority values you can choose are 0, 4096, 8192, 12288, ... 61440. If you set the STP priority to 0, that switch will be the root bridge.

NB:It is not best practice to set the STP priority to 0 on more than one switch. Setting the STP priority to 0 on more than one switch would basically tell STP to use those switches as the root bridge. However, you can only have one root bridge in any given Layer 2 spanning tree, so STP will simply choose the switch with the lowest MAC address among the switches with STP priority set to 0.



Assigning STP port types

The second step in the STP operation flow is assigning port types to every port in the network. STP assigns port types according the cost of each path between the root switch and a nonroot switch.

STP path cost

Switches are typically interconnected using redundant links to increase resilience, should one of the links fail. Each of these links provides a connection path. Each path has a certain cost associated with it. The cost is defined according to the bandwidth of the link. Table 4-1 lists the costs associated with each Ethernet bandwidth.

Table 4-1	STP Path Cost for Each Ethernet Bandwidth
Bandwidth	*STP Cost*
10 Mbps	100
100 Mbps	19
1 Gbps	4
10 Gbps	2

STP prefers paths with lower costs

STP will designate the port that connects to the 1-Gbps link to be root port, the port that connects to the root bridge. The port connecting to the slower 100-Mbps link will be either a designated port or a blocking port. A designated port is a port that forwards data-link frames. A blocking port is a port that does not forward data-link frames.

Setting root ports

After switches have elected a root bridge, nonroot switches assign one of their ports to be their root port. The STP root port is the port that connects a switch to the STP root bridge. The port that is elected to be the root port is either

✦ A port that connects the nonroot switch directly to the root bridge
✦ A port that connects to the least expensive path (STP cost) to the root bridge

Switches communicate between themselves using STP to calculate the cost of each path to the root bridge. Each switch adds the cost of its own path to the cost received from the neighboring switches to determine the total cost of a given path to the root. After the costs of all paths to the root have been calculated, each switch assigns the root port to the port connecting to the path with the lowest cost.

Setting designated ports

After each switch assigned one of its ports to be the root port, connecting to the root bridge, the remaining ports are either *designated* or *blocking:*

✦ An STP designated port is a port that forwards data-link frames.
Each LAN segment needs to have a designated port, a port that forwards traffic in and out of that LAN segment.
✦ An STP blocking port is a port that does not forward data-link frames.

Setting blocking ports

Earlier you read that each LAN segment needs to have a designated port, a port that forwards traffic in and out of that LAN segment. The port at the other end of the link becomes a blocking port. An STP blocking port is a port that does not forward data-link frames, thereby closing the loop. A blocking port still receives data frames, but it sends no data frames out on the link. In other words, the link becomes one-way, from the designated port to the blocking port.

Achieving STP convergence

After switches have assigned a forwarding type or a blocking type to all ports that interconnect them, STP achieves a stable, loop-free LAN. This is a *converged network.* Convergence is the result of assigning port types to eliminate loops in the LAN.

Introducing bridge protocol data units (BPDUs)

Switches communicate by sending each other bridge protocol data units(BPDUs).
A BPDU is a special-purpose data-link frame that is multicast every 2 seconds. BPDUs contain information about STP path costs, bridge IDs, port IDs, and some other parameters that help switches to elect the root bridge and to decide how to assign port types to their respective ports.

Topology change notification (TCN) BPDUs

What if the topology of the LAN changes? What if you add or remove a link between two switches, or if you add or remove a switch? In that case, the switches involved send a topology change notification (TCN) BPDU. A TCN is a specialized BPDU that informs every switch in the LAN that the

topology has changed. At that point, switches need to go through the convergence process again: They need to reelect a root if the root bridge was affected by the topology change. Next, they need to decide how they set up their ports. Finally, they achieve a converged state.

STP port states

STP Port State	Send/Receive BPDUs	Frame forwarding (regular traffic)	MAC address learning	Stable/ Transitional
Blocking	NO/YES	NO	NO	Stable
Listening	YES/YES	NO	NO	Transitional
Learning	YES/YES	NO	YES	Transitional
Forwarding	YES/YES	YES	YES	Stable
Disabled	NO/NO	NO	NO	Stable

Ports on switches running STP are in one of the following states:
Blocking: Ports start in the blocking state. A blocking port forwards no data-link frames. It listens to what's happening in the LAN: It receives and analyzes BPDUs. Having all ports blocking when a switch is powered up prevents the switch from creating and using transmission loops while STP is converging. A port that is an endpoint to an interswitch link stays in the blocking state unless it becomes a root port or a designated port during the STP convergence process. A port can also transition to the blocking state if the switch receives a topology change notification (TCN) BPDU. Whenever the network topology changes, STP blocks all ports until the STP convergence process is restarted. A delay occurs before the STP convergence is restarted. Whenever a switch or a link fails or becomes unavailable, switches connected to the failing switch or link wait for 20 seconds before they start the STP convergence process. This 20-second wait is called the max age timer. You can view and change the max age timer using Cisco IOS commands.
Blocking state duration: 20 seconds (max age timer)

Listening: Next, ports transition to the listening state. A listening port listens to BPDUs to prepare to transmit frames. A port in the listening state would only send and receive BPDUs without populating the MAC address table of the switch. This is the state in which switches communicate using BPDUs to assign a port type to the listening port. A port stays in the listening state for 15 seconds by default. The listening time and the learning time make up the forward delay timer. You can view and change the forward delay timer using Cisco IOS commands.
Listening state duration: 15 seconds (forward delay timer - part I)

Learning: Ports transition to the learning state. A learning port listens to BPDUs and populates its MAC address table. The purpose of the learning state is to allow the switch to gather information about the MAC address reachable on each port. A port in learning mode sends no data frames. A port stays in the learning state for 15 seconds by default. The listening time and the learning time make up the forward delay timer. You can view and change the forward delay timer using Cisco IOS commands.

Learning state duration: 15 seconds (forward delay timer - part 2)

Forwarding: At this point, the port has become a root port, a designated port, or a blocking port. If the port is either root or designated, it transitions to the forwarding state. A port in the forwarding state can send and receive data frames.
Disabled: A disabled port is basically a port that has been shut down manually by the switch administrator. Disabled ports do not participate in the network: They do not go through the convergence process, and they do not send or receive frames. They are basically turned off.

Options for STP

PortFast
The Cisco PortFast port option is used on ports that do not need to participate in STP. These are typically ports that do not interconnect switches, bridges, or hubs. For example, a switch port that connects directly to a host device using a single link does not need to participate in STP because there are no chances that that port will ever create a switching loop.

BPDUGuard
The BPDUGuard Cisco option is used in conjunction with the PortFast option on access layer switches. The PortFast option is dangerous if it is enabled by mistake on a port that interconnects switches: Enabling PortFast turns off STP on that port. This can potentially create a switching loop. The BPDUGuard option monitors the frames received on the PortFast-enabled port.

BPDUFilter
The BPDUFilter option does not allow BPDU frames to enter or exit a PortFast-enabled port. Without BPDUFilter enabled, a port that is PortFast enabled still receives BPDU frames. BPDUs are only useful in the context of STP. So, it makes sense to enable BPDUFilter to fence off the BPDUs on a port that is PortFast enabled because that port does not participate in STP.

Best practice is to use at least one of these options along with PortFast for ports that do not need to participate in STP.

Rapid Spanning Tree Protocol (RSTP)

Layer 2 switches need to avoid switching loops. However, STP convergence delays are a problem. Following STP best practices increases the chances to have a quick STP convergence after a topology change, but you have no guarantee that it will be any quicker than 50 seconds**.**

Considering that BPDUs are sent every 2 seconds, STP switches wait for 10
BPDUs before they start the STP recalculation process. RSTP reduces this 20-second delay. RSTP waits for just three BPDUs before it starts the STP convergence process. This considerably speeds the STP convergence process: a 6 max age timer delay in RSTP instead of the 20-second
max age timer delay in STP.

Alternate port and backup port

Alternate port

An *alternate port* is a port that becomes the root port in the event that the root port or the link starting at the root port fails or becomes unavailable.

Backup port
A *backup port* is a port that becomes the designated port in the event that the designated port or the link starting at the designated port fails or becomes unavailable.

For further reading important for future upscaling:

EtherChannel

Why would you have redundant links between switches if STP effectively shuts them down? Why would you spend extra money to have redundant links if you cannot use them? The additional links are only used for failover. Wouldn't it be nice if you could use these extra links to send data, thereby increasing the throughput of your LAN? That's exactly what EtherChannel does.

EtherChannel allows you to group several physical links into a single logical link. This is also called *port trunking* because you put several ports in a logical trunk.(link aggregation)

EtherChannel and STP are friends

It is best practice to enable EtherChannel on redundant interswitch links when using STP.

So, STP is happy because only one (logical trunk) port exists between the switches: You have no danger of creating a switching loop, so you don't need to block any ports.

Enabling EtherChannel on your redundant links provides three main advantages:

✦ By grouping up to eight physical ports into a single logical port trunk, you add the bandwidth of each port to provide an *increased bandwidth* in the port trunk. For example, if you group eight Fast Ethernet ports into a port trunk, your total bandwidth in the port trunk would be 8 × 100 Mbps = 800 Mbps.

✦ EtherChannel uses *load-balancing* algorithms to spread the network traffic across all ports in the port trunk. If one of the ports becomes overloaded, traffic is distributed across the remaining ports.

✦ EtherChannel has built-in *fault tolerance:* If one port or link fails, EtherChannel sends traffic across the remaining ports in the trunk.

Link Aggregation Control Protocol (LACP)
Port Aggregation Protocol (PAgP)