Extra topics in Networks

STP

- **Spanning Tree Protocol** builds a loop-free logical topology of Ethernet networks.
    - Connections between Ethernet switches typically contain loops for redundancy, STP disables (blocks) some ports to prevent **problems caused by switching loops** and re-enables them upon need.
        - **Problems caused by switching loops:**
            - **Broadcast storm**: broadcast messages being retransmitted over and over along the loop, eventually congesting and taking down the network).
            - **Unstable MAC addresses:** ARP tables keep updating because information is being received from different ports.
            - **Duplicate frames** can reach the same host from different paths.
    - **Which port(s) to disable?**
        - This is determined by the spanning tree algorithm discussed below.
        - Overall the spanning tree created by disabling ports is the one that have the minimum path cost (ties are broken by a well-defined criteria).
    - **Port states:** a port in an STP-enabled switch can be in one of the following states:
        - **Disabled:** shutdown.
        - **Blocking:** receiving traffic but ignoring them.
        - **Listening:** not forwarding traffic, not learning MAC addresses (typically stays there for 15 seconds before moving to Learning state)
        - **Learning:** not forwarding traffic, but learning MAC addresses (typically stays there for 15 seconds before moving to Forwarding state state)
        - **Forwarding:** forwarding traffic and learning MAC addresses as normal.

    - **How it works:**
        1. Switches elect a root bridge (king of switches).
            - Switches exchange Ethernet frames containing their BPDUs as payload, the switch with the lowest Bridge ID wins.
                - **Bridge Protocol Data Unit (BPDU)** is the data format used by switches for exchanging STP information.

                - Bridge ID = priority (4 bits) + locally assigned system ID extension such as VLAN ID (12 bits) + ID [MAC address] (48 bits); the default bridge priority is 32768.
                - Port ID = priority (4 bits) + ID (Interface number) (12 bits); the default port priority is 128.
        2. Root bridge interfaces are set to be in the **Forwarding** state.
        3. Each non-root switch selects one **Root Port (RP)**.
            - RPs are the ones that have the smallest root cost (path cost to reach the root bridge)

- **Root cost from switch X to the root bridge = the sum of all outgoing port costs**
- **Port costs:**

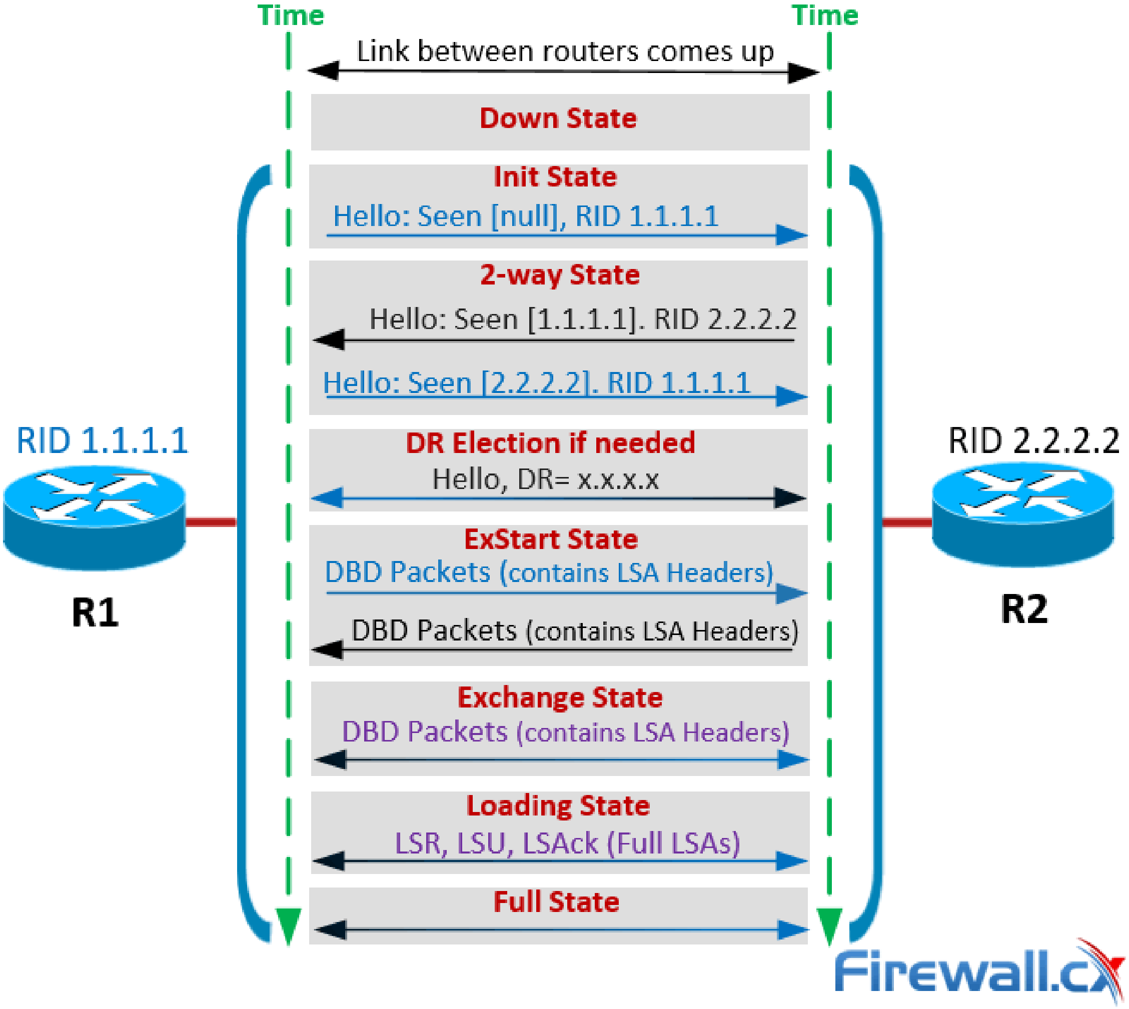
Port Speed	Original	New Cost
10 Mbps	100	2,000,000
100 Mbps	19	200,000
1 Gbps	4	20,000
10 Gbps	2	2000
100 Gbps	N/A	200
1 Tbps	N/A	20

4. Each remaining link chooses one **Designated Port (DP)**.
    - DP is a non-root port that is permitted to forward traffic, it's the one that minimizes path cost to the root bridge.
    - If both sides of a link are non-RPs, only one of them will be a DP, and the other will be set to a **blocked** state.
5. All non-RPs, non-DPs are blocked.

- **STP Types**
    - **Original STP (802.1D)**: converges slowly.
    - **Rapid STP (RSTP) 802.1w:** improved version of STP with faster convergence.
    - **PVST+ and Rapid PVST+** are Cisco's proprietary improvement of the standard STP and RSTP, respectively.
        - They added the "per-vlan" feature that allows creating different STP configuration for each VLAN.
- **PortFast** and **BPDU Guard** are STP extensions that are typically used together to allow fast transition between blocking and forwarding states while ensuring that loops don't arise.

OSPF

- Open Shortest Path First is a widely used and supported **dynamic**, **global**, **interior gateway (inter-AS)**, **link-state** and **load-insensitive** routing protocol.
    - **Dynamic:** routing information are learned from the network and not configured statically.
    - **Global:** all router know the same information about all links in the autonomous system.
    - **Interior gateway:** used to exchange routing and reachability information between routers in the same **Autonomous System (AS)**.
    - **Link-state:** each node send information to all other nodes regarding the state of it's neighboring links.relay
    - **Load-insensitive:** the best path chosen doesn't depend on the current traffic load in the network, a congested path can be chosen as the best one if it has the minimum cost.
- **How it works:**


Time
Time
Link between routers comes up
Down State
Init State
Hello: Seen [null], RID 1.1.1.1
2-way State
Hello: Seen [1.1.1.1]. RID 2.2.2.2
Hello: Seen [2.2.2.2]. RID 1.1.1.1
DR Election if needed
Hello, DR= x.x.x.x
ExStart State
DBD Packets (contains LSA Headers)
DBD Packets (contains LSA Headers)
Exchange State
DBD Packets (contains LSA Headers)
Loading State
LSR, LSU, LSAck (Full LSAs)
Full State
RID 1.1.1.1
R1
RID 2.2.2.2
R2
Firewall.cx

1. OSPF routers agree to form a neighbor relationship.
    - **(Down state)** Each router chooses a RID: static -> highest loopback interface IP -> highest non-loopback interface IP.
    - **(Attempt/Init state)** Each router sends a HELLO message to its neighbors, containing it's RID and known neighbors.
        - If receiver is already neighbor with sender, it just resets dead interval.
            - If dead interval passed with no received hello from neighbor, it's assumed to be dead and set to down.
        - Else, it starts building neighbor relationship with sender.
            1. Check that attributes match (e.g., area ID)
            2. Add sender RID to its list of known neighbors.
            3. Enters **(2-way state)**.
    - **(DR election)** on broadcast networks, routers will only continue form full-neighbor relationship with Designated Router (DR).
        - DR works as the central point of trusted LSA updates, it receives link state updates and share them with the rest of the network.
        - DR is selected at this stage (manually assigned priority -> highest RID) and announced to everyone.
        - A Backup DR (BDR) is also elected to serve in case the DR is down.
    - **(ExStart):** master router (higher RID) is selected, it generates an initial sequence number to be used for exchanging LSDB data with slave routers.
    - **(Exchange):** LSDBs are exchanged between routers.
        - Each router broadcasts **Link-State Advertisements (LSAs)** packets containing information about each subnet it's part of.

        - LSAs are re-broadcasted whenever there is a change, or just every specified interval (to ensure paths are updated).
        - Received information is stored in each router's **Link-State Database (LSDB)**
    - **(Loading):** now each router knows LSDBs of other routers, it compares it with its own LSDB
        - If a neighbor **B** knows about an interface which **A** doesn't know about, **A** will ask **B** for this information (LS Request), **B** will reply with data (LS Update), **A** will then send LSAcknowledge for updates.
        - Process continues until every router knows about every subnet in the AS and LSDBs are synchronized.
    - **(Full)** routers now have formed full neighbor relationship with DR and BDR and can load upcoming updates.
2. Each router chooses the best routes based on information exchanged from LSDBs (runs Dijkstra single-source shortest path algorithm locally)
    - Each link is assigned a cost based on its bandwidth, costs are used as weights while applying Dijkstra's algorithm.
    - Serial interface: 64, Ethernet: 10, FastEthernet: 1, anything faster: 1.

EIGRP

- Enhanced Interior Gateway Routing Protocol was a Cisco proprietary routing protocol before being published as an open standard.
- It combines ideas from link-state and distance vector routing protocols, but can be considered a distance vector protocol in the sense that routers only advertise link state information to their immediate neighbors only.
- Similar to OSPF, routers must first form neighbor relationship, they exchange routing information, which each router use to determine the best routes.
- EIGRP uses Reliable Transport Protocol (RTP) for information exchange.
- Hello messages are send to a multicast address of 224.0.0.10

IS-IS

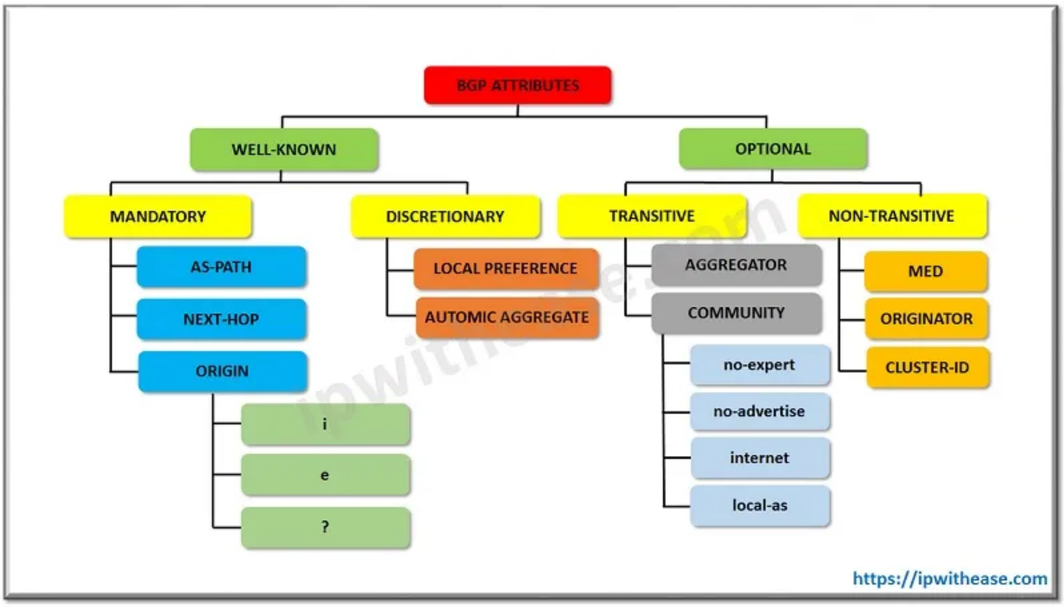
- Intermediate System to Intermediate System is a historic standard for Inter-AS routing.
- Like OSPF, it's a link-state protocol that floods link-state information and use Dijkstra to find shortest path.
- It uses **TCP port 2024** for transporting information.

RIP

- Routing Information Protocol is an old Inter-AS, distance vector routing protocol.
- It uses **hop count** as the routing metric and calculates shortest path using Bellman-Ford.
- It is easy to configure, but has slower convergence time compared to OSPF, ISIS, or EIGRP.
- It **prevents routing loops** by limiting hop count to 15.
- It **prevents propagation of incorrect information** using split horizon, route poisoning, and holddown mechanisms.
- It uses **UDP port 520** for transporting information.

BGP

- Border Gateway Protocol is the **dynamic, decentralized, exterior gateway (inter-AS), distance vector, and load-insensitive** protocol used mainly by the Internet to exchange **routing and reachability** information among ISPs.
    - **Decentralized**: routing information is not stored in one place, no single point of control/failure on routing tables.
    - **Distance vector:** each router sends prefix reachability information to it's neighbors, the best route selection algorithm is based on Bellman-Ford.
    - **Reachability information:** subnet X exists.
    - **Routing information:** a possible path to get to subnet X is through AS Y.
        - When a router knows about multiple paths to a subnet (prefix), it runs the route selection algorithm to choose the lest-cost one.
- **BGP connections**
    - **External BGP connection (eBGP):** a TCP connection on port 179 between two gateway routers belonging to different ASs.
    - **Internal BGP connection (iBGP):** a TCP connection on port 179 between two routers belonging to the same AS.
- **BGP Attributes:** a BGP advertisement message contains many attributes, the well-known mandatory ones are:
    - **ORIGIN:** can be
        - `i`nternal: the route originated from an IGP (e.g., OSPF, RIP, EIGRP)
        - `e`xternal: the route originated from an EGP (i.e., another BGP router)
        - `?` (incomplete): the route originated statically, and was distributed by an IGP.
    - **AS_PATH:** the list of ASNs through which the advertisement has passed.
        - Each AS has a unique number associated with it by the IANA.
        - AS_PATH is also used for loop prevention, if an AS finds its ASN in AS_PATH, it rejects the advertisement.
    - **NEXT_HOP:** the IP address of the gateway router interface of the first AS in AS_PATH.
        - Whenever a router receives an advertisement for a path it already knows a route for, it will update its routing table if that path has a lower cost according to the locally-run route selection algorithm.



- **Transitivity of BGP attributes**

    - A transitive BGP attribute is allowed to be sent to other peers, well-known attributes are transitive.
    - Transitive attributes will propagate through the ASs, while non-transitive ones will only be sent to neighbors.
- **BGP elimination rules**: to determine the route to subnet "X" from AS "A".
    1. From the list of possible routes to X, keep only the ones with the **highest local preference**.
    2. From the remaining routes, keep only the ones with the **smallest AS_PATH length**.
    3. For the remaining routes, use **hot-potato routing** to choose the best one.
        - **Hot-potato routing:**
            - Get that hot-potato (packet) outside of our AS (give it to the NEXT_HOP) as quickly as possible.
            - The route chosen (from among all possible routes) is that route with the least cost to the NEXT_HOP router beginning that route.
            - **Example**: a packet from 1b is destined to subnet X, 1b will ask its intra-AS routing algorithm, which has a lower cost, the path to 2a or to 3d.
    4. If more that one route still remains, use BGP identifiers to choose the best one.
        - **Equal-Cost Multi-Path (ECMP) routing** is the strategy that allows routing to a single destination using multiple equal-cost paths, with the ability to load-balance between them, this is however difficult to deploy in practice.
- **Route Reflector (RR)**
    - For iBGP connections, all peers must be connected to each (i.e., an AS must be connected in a full-mesh topology), this doesn't scale.
    - **Route Reflectors** reduce the number of required connection using an approach similar to DR and BDR election in OSPF.
    - A RR is a router that acts as a focal point in for iBGP sessions, it propagates iBGP information to all routers in the AS without the need to have a full mesh topology.
- **AS Path perpending**
    - A trick to reduce the priority of going through a certain AS, by adding the same ASN multiple times to make the AS_PATH longer.
- **BGP Confederation**
    - A group of ASs that appears to BGP as a single AS with one common ASN.
- **BGP Split Horizon**
    - Never send routing information back in the direction (same interface) from which it was received.
    - A router that receives routing information from an iBGP peer should not forward this information to other iBGP peers
    - It prevents routing loops in distance vector protocols.

First Hop Redundancy Protocols (FHRP)

- A family of protocols used to support redundancy in layer 3 routers without having to have multiple configured gateway addresses on hosts.

- The idea is generally to have a group of redundant routers that share the same (virtual) IP (the one configured as gateway in hosts).
- Implementations include HSRP, VRRP, and GLBP
  - Host-Standby Router Protocol (Cisco proprietary)
  - Virtual Router Redundancy Protocol (RFC 5798)
  - Gateway Load Balancing Protocol (Cisco proprietary)
- **VRRP**
  - The group of redundant routers share the same (virtual) IP and (virtual) MAC address format (0000.5e00.01XX) where XX is replaced with group ID.
  - Traffic intended for gateway goes through **the master** router, or **a backup** one if master fails.
  - The master router is chosen to be the physical owner of the virtual IP, if there is no such router -> highest priority -> highest IP address.
  - The master router multicasts advertisements (224.0.0.18) every second to all devices indicating that it's alive, if master down timer (3 seconds) passed, the backup router takes over and replaces master.
  - When master is back, it takes over again.

Focus: DHCP

- Dynamic Host Configuration Protocol automatically configures new hosts joining the network, or an existing host that was down and back up.
- It assigns values to the host such as the IP address, subnet mask, gateway address, DNS server address, and/or the time server.

- It typically involves 4 steps, the first two may be omitted if the host is refreshing it's IP address rather that requesting a new one.
  - **DHCPDiscover:** the client broadcasts a message discovering the DHCP server (it can be the same as gateway router)
    - For IPv4, source address will be 0.0.0.0 and destination is 255.255.255.255
    - A certain transaction ID is also sent for the server to use in replies.
  - **DHCPOffer:** the server broadcasts an offer message offering a certain IP to the host, the message contain the same transaction ID, offer depends on the DHCP mode.
    - **Automatic DHCP:** server offers clients any address from the pool of available addresses, but once an address is assigned, host takes it permanently (unless it releases the address voluntarily)
    - **Dynamic DHCP:** server offers clients any address from the pool of available addresses, but addresses assigned has a lease (expiration) time, after which the host should refresh the address if it wants to keep using it.
      - This is the most common DHCP type.
    - **Manual:** the administrator maintains a static table that assigns the same IP address to the same host (characterized by its MAC address).
  - **DHCPRequest:** the client accepts the first offer it gets and can either
    - **Send the DHCPRequest directly to the server (unicast)** to save bandwidth, the host should know the IP address of the DHCP server (from the offer).

        - **Broadcast DHCPRequest** telling the server that it accepted a certain offer (attaches the offer transaction ID incremented).
            - DHCPRequest broadcasting is only valid if the DHCP server is on the same network since routers do not forward broadcast messages.
            - To solve this problem **DHCP relay agent** (discussed below) is used.
    - **DHCPACK:** the server broadcasts (or unicasts) an acknowledgement to the client confirming that it can now use the offered IP address.
        - This is needed to ensure the DHCPRequest was not lost and the server knows about the allocated IP (not available to other hosts).
        - If a client doesn't need the IP address anymore, it can send a **DHCPRequest** message to the server.
- **DHCP Relay agent**
    - Allows using a single DHCP server to manage a big network (with different subnets).
    - It acts on behalf of the DHCP client by taking the broadcast messages sent by the client and unicast them to the server and broadcasts the returned server replies.
- **Rogue DHCP server**
    - A Man-In-The-Middle attack where a malicious host on the same subnet as the DHCP client deceives the client by acting as the DHCP server and replying to DHCPDiscover, offering it's own IP address as the default gateway.
    - This will allow the attacker to intercept all requests sent by the victim.

SNMP

- The standard protocol for collecting and organizing information about managed (and monitored) IP hosts.
- The management data is organized in a management information base (MIB) which describe the system status and configuration.
- These variables can then be remotely queried (and, in some circumstances, manipulated) by managing applications.
- Net-SNMP is a suite of software for using and deploying the SNMP protocol

EtherChannel

- Link aggregation technology that allows **grouping two or more** (typically up to 8) **physical Ethernet links** (connected in parallel between two devices) **to act as one logical interface**
- EtherChannel make use of the combined bandwidth (while maintaining load-balance among links).
- If a link is down, others will be utilized, no STP reconfiguration when link is replaced.
- Configuring EtherChannel can be done in 3 modes: **static**, **PAgP** (Port Aggregation Protocol), or **LACP** (Link Aggregation Control Protocol).
    - A PAgP or LACP port can be in on of two modes: (desirable/active, auto/passive), the active mode allows the port to initiate auto-config negotiations with the other active/passive port, passive-passive configuration won't allow EtherChannel to work.
- For EtherChannel to work, ports has to have the same duplex, speed, type (access/trunk), STP interface settings.
    - If using access ports (connected to hosts), they have to be in the same VLAN.

  - If using trunk ports (connected to another switch), they have to have the same native VLAN and same allowed VLANs.

VTP

- **VLAN Trunking Protocol** is used to synchronize VLAN configurations among Ethernet switches.
- Each device has a VLAN configuration database stamped with a revision number.
- A **server** switch propagates database information to other switches.
- A **client** switch receives information from server (either directly or propagated from another switch) and updates its configuration if it has a lower revision number.
- A **transparent** switch doesn't participate in the "game" but forwards traffic to other participants.