yoinked from https://github.com/Nicicalu/OST-CN2-Spick

OSPF (IGP)

Equal Cost Multipath (ECMP) load balancing, fast convergence, widely used. Uses IP Port 89 on L3. Uses `224.0.0.5` (all routers), `224.0.0.6` (all DR,BDR)
DR Election: Highest IF prio → if tie, highest Router ID. Priority 0 = ineligible for DR/BDR. No preemption. BDR = same rules, second best candidate
Virtual links: Cannot go through more than one area, can only run through standard non-backbone areas. (eg. not over stubby areas)

ROUTER OPERATION

- Establish neighbor adjacencies and exchange LSAs
- Build the Link-State Database (LSDB)
- Run Dijkstra's SPF algorithm:
  - **Intra-area change:** SPF recalculation
  - **Inter-area change:** No SPF needed — ABR handles updates
- Build the routing table from SPF results

ROUTE TYPES

O: Intra-area: Local area → Local area. Type 1 and 2 LSAs
O IA: Inter-area: Local area → Neighboring area. Via LSA 3 through backbone
O E1 or O E2: External: Local area → External (eg. BGP), from ASBR
O E1: Cost = internal + external metrics
O E2: Cost = external metrics. Default external route for OSPF
Cost calculation: Reference Bandwidth(def=100 Mbps)/Interface Bandwidth

AREAS

AS split into **areas** with 32-bit Area ID (e.g., `0.0.0.0` = Area 0)
Backbone Area (Area 0): Core of the OSPF domain, must connect to all other areas (directly/virtual links), must be contiguous (no disjointed segments), should not contain end-user networks, summarizes topology of other areas
Non-Backbone Areas: Connect end users and local resources, All inter-area traffic must transit the backbone

ROUTER TYPES

Area Border Router (ABR): Connects two or more OSPF areas, must have 1 interface in backbone. 1 OSPF DB per Area
Internal Router (IR): All interfaces belong to the same area (non-backbone)
Backbone Router (BR): At least one interface in Area 0, Includes ABRs and routers internal to the backbone
AS Boundary Router (ASBR): Connects to external AS/Network (e.g., BGP), Advertises external routes into OSPF, can be in backbone or non-backbone

DESIGN RULES

Rule 1: Backbone (Area 0) must be contiguous — no partitions allowed
Rule 2: Every non-backbone area must connect to Area 0

LINK STATE ADVERTISEMENT (LSA) TYPES

(1) Router-LSA: ***All routers*** list directly connected links (so all outgoing interfaces with state and cost) (**intra-area**)
(2) Network-LSA: ***DR*** lists all routers and DR in broadcast/multi-access network (**intra-area**)
(3) Summary-LSA: ***ABR*** advertises networks from other areas, flooded in all the areas that are not "totally stubby" (**inter-area**)
(4) ASBR-Summary: ***ABR*** advertises path to ASBRs. ASBR sends Type 1 + external bit, ABR converts to Type 4 and floods (**inter-area**)
(5) AS-External-LSA: ***ASBR/ABR*** advertises external routes (e.g., BGP); flooded to all non-stub areas → only normal areas (**inter-area**)
(7) NSSA External: Like Type 5 but inside NSSA. Converted to Type 7 by ABR. Shown in routing table as **(O N1 or O N2)**

AREA TYPES (ALLOWED LSA TYPES)

Stubby = Blocks external LSA [No 5], Totally = Blocks inter-area LSA [No 3/4]
Default route (`0.0.0.0`) from ABR [DR ABR], Cannot contain ASBR [No ASBR]
Standard (1,2,3,4,5): Normal
Stub Area (1,2,3): [No 4/5] [DR ABR] [No ASBR]
Totally Stubby Area (1,2): [No 3/4/5] [DR ABR] [No ASBR]
Not-So-Stubby Area (1,2,3,7): [No 4/5], no DR, allows 1 ASBR [LSA 7 → 5]
Totally NSSA (1,2,7): [No 3/4/5] [DR ABR] [LSA 7 → 5]

PACKET TYPES

(1) Hello: Usually Multicast. Discover, maintain, and verify neighbors. Forms adjacencies. Election of DR,BDR in broadcast networks. Contains network mask (sending routers' IF), Hello interval (p2p,broadcast: def=10s), Options, Priority (election), Router dead interval (def=40s), DR/BDR IP, Neighbors.
(2) Database Description (DBD/DD): Unicast. Exchange summaries of LSAs during adjacency formation (headers only). Contains IF Max. MTU, Options, I/M/MS bits (Initial, More, Master-slave bit), DD Seq. Num., LSA Header
(3) Link State Request (LSR): Unicast. Request specific LSAs. Contains Link State Type (router/net), Link State ID, Advertising Router (sender address)
(4) Link State Update (LSU): Multi/Unicast. Flood new/updated LSAs, Answer to LSR (unicast), Contains Nr. of LSAs, full LSAs information.
(5) Link State Acknowledgment (LSAck): Multi/Unicast. **Explicit ack.** Confirms receipt of LSAs for reliable flooding. Acks may be grouped together.

SUB-PROTOCOLS

Hello Protocol: Neighbor discovery. Parameter negotiation. Maintain logical adjacencies on P2P, P2MP, virtual links. Elect DR/BDR on broadcast and NBMA networks. Continuously sends hello packets to maintain bidirectional connectivity; failure to receive = neighbor down (RouterDeadInterval)
Database Sync Protocol: Sync LSDB using DBD packets.

NEIGHBOR STATES

Down: Initial state, no Hello packets exchanged or after RouterDeadInterval
Init: Received Hello packet from neighbor but without its own Router ID yet
2-Way: Seen its own Router ID. DR/BRD election is taking place in NBMA
ExStart: (DBD) Bi-dir comm; highest Router-ID=master. Determine init Seq Nr
Exchange: Exchange of DBD packets (LSA headers).
Loading: Missing LSAs are requested. (LSR/LSU)
Full: Databases fully synchronized. Normal operating state

OSPF ROUTING AND ECMP

- Each router runs Dijkstra per area; link cost = metric from LSAs (1–65535)
- Perfer more specific match (CIDR), else O > O IA > E1 > N1 > E2 > N2
- Routes added to RIB/FIB based on computed next hops
- ECMP: Modified Dijkstra supports Equal-Cost MultiPath if multiple paths have same cost → routes added with multiple next-hops for load balancing

IS-IS (IGP)

Widely used by ISPs, Fast convergence, ECMP, Extendable
ES: End-host devices are called End Systems (ES)
IS: Routers are called Intermediate Systems (IS)

CONNECTIONLESS NETWORK SERVICES (CLNS)

CLNS: ISO Layer 3 datagram service; supports **CLNP**, **ES-IS**, **IS-IS**.
CLNP: Connectionless Network Protocol, similar to IP, used in ISO stack
ES-IS: host ↔ router discovery protocol
IS-IS: router ↔ router. Link-state routing protocol
Integrated IS-IS: Allows IP routing with IS-IS

NETWORK SERVICE ACCESS POINT (NSAP)

Variabale in length (up to 20 bytes), Identifies node, not IF. Example from IP: `192.168.1.24` → `49.0001.1921.6800.1024.00`

`49.0011.0000.0000.0003.00`
Area ID (11) System ID (R3) NSEL

Network Entity Title (NET): Unique router identifier. NSEL=0,Starts with 49
Authority & Format Identifier (AFI): NSAP format+authority that assigned it
Initial Domain Identifier (IDI): Variable len, identifies administrative domain
Domain Specific Part Format Identifier (DFI): Specifies format
Domain Specific Part (DSP): Var len, contains hierarchical structure of addr
NSEL (N-Selector): always `00` for routers (1b)
System ID: Unique per router (often based on lo-IP/MAC) (6b)
Area ID: Used for routing hierarchy (like OSPF areas) (AFI+IDI+DSP)

PACKET TYPES (ALL IN L1 OR L2)

All PDUs contain Header with shared common fields + specific routing-related information (Type, Length and Value (TLV)) used to extend protocol
IIH (Hello): Build,maintain adjacencies. Includes system ID, holding time, prio
  Functions: discover, build, maintain
  Interval: 10s (default); DIS sends every 3.3s on LANs
  Multiplier: Missed Hello limit → Holdtime = Interval × Multiplier(def=3)
  Multicast: [L1 `01-80-C2-00-00-14`] [L2 `01-80-C2-00-00-15`]
LSP (Link State PDU): Contains topology info including prefixes with costs; unicast on p2p or flooded through the area (OSPF: LSA Type 1)
CSNP (Complete SNP): Sent by DIS; lists all known LSPs (OSPF: DBD)
PSNP (Partial SNP): Used to request missing or acknowledge received LSPs

ROUTER TYPES

L1 Router: Within one area; all have same LSPDB. No inter-area routing
L2 Router: Backbone router; all have same LSPDB. Routes between areas
L1/L2 Router: Acts as both; separates LSPDBs, redistributes between levels

ADJACENCIES

Backbone must be contiguous: L1 ↔ L1-L2 ↔ L2
L1: Are addresssses match unless configured otherwise
L2: Alongside L1, unless router is L1 only, areas must not match

POINT-TO-POINT (NO DIS)

Initialized by receipt of ISHs, exchange p2p IIHs padded to MTU
CSNP: Sent once at adjacency startup
LSP: Advertises topology changes (link-state info)
PSNP: Acknowledges received LSPs or requests missing ones

BROADCAST/MULTI-ACCESS (DIS REQUIRED)

Not triggered by ISHs, instead broadcast IIHs with all neighbors MAC's
Designated Intermediary System (DIS): Create, update pseudonode LSPs. Flood LSPs. Simlar to DR in OSPF. No backup DIS.
DIS Election: Highest priority (0–127) (Cisco def=64) > Highest SNPA (MAC)
Preemption: Enabled → higher prio router automatically takes over DIS Role
Pseudonode: Carried out by DIS. Multiaccess links. Separate for L1,L2
Passive interfaces: Advertise network prefixes without adjacency forming

PATH SELECTION

Default interface metric = 10. **Lower Metric better:** L1 intra-area > L2 intra-area > Leaked L2 → L1 (internal metric) > L1 external (external metric) > L2 external (external metric) > Leaked L2 → L1 (external metric)

LEVEL 1 ROUTING

- Intra-area routing only (L1 Area like OSPF Totally Stubby Area)
- L1 routers install a **default route** to nearest L1-L2 for inter-area traffic
- **L1-L2:** Do **not** advertise L2 routes into L1 area (unless route leaking active)
- **L1-L2:** Set **Attached bit** to signal L2 connectivity to backbone
- **Distribution Bit:** Set to 1 on L2 → L1; blocks re-advertisement L1 → L2.
- **Route-Leaking** injects a more specific route into L1 to improve routing

LEVEL 2 ROUTING

- Routing between areas (inter-area)
- L1-L2 routers inject L1 routes into L2 topology
- L1 routes are redistributed into L2 with L1 metric preserved in L2 LSP

IS-IS VS OSPF

Feature	*IS-IS*	*OSPF*
Layer	L2 (CLNS)	L3 (IP, proto 89)
Encapsulation	No IP, uses TLVs	IP packets
Hello Type	IIH	Hello packet
Area Model	L1/L2	Backbone + Areas
Metric	Cost (default 10)	Cost (bandwidth)
Router ID	System ID (6B)	32-bit Router ID
Adj. Types	L1, L2, L1/L2	DR/BDR, P2P
LSDB	Per level (L1/L2)	Per area
Scaling	Large-scale ISP core	Enterprise/campus
Routing Info	TLVs (flexible)	Fixed LSA types
Type	Link-State	Link-State
Shortest Path calculation	Dijkstra	Dijkstra

BGP (EGP)

Path-vector routing protocol (doesn't contain full topology). Stable, flexible, scalable. Based on TCP. Loop prevention using AS_PATH

IBGP

Routers in same AS, AD=200, more trusted (lower security overhead). AS-Path and Next-hop not modified

SCALABILITY

- iBGP does not re-advertise routes between iBGP peers → full mesh required (cause no loop prevention)
- Session count = $^{n(n-1)}/_{2}$ (e.g., 5 routers = 10 sessions, 10 = 45)

ROUTE REFLECTORS (RR)

Solves iBGP full-mesh scaling by allowing selective route reflection. Clients only peer with RR; unaware they're clients. Only the RR needs special config.
From non-client: RR advertises to **clients only**
From client: RR advertises to **all** (clients + non-clients)
From eBGP peer: RR advertises to **all** (clients + non-clients)

EBGP

Routers in different ASes, AD=20, stricter policy enforcement. Prepends ASN to AS-Path, modifies next-hop IP. Fails if own ASN in AS-Path. TTL=1
AS-Override: Allows reuse of same ASN across different customer sites (e.g., Swisscom); rewrites ASN to avoid loop detection

CONFIG

next-hop-self fixing iBGP: neighbor (neighbor IP) next-hop-self (overriding)

AUTONOMOUS SYSTEM NUMBERS (ASN)

- Unique ID for each AS; required for Internet routing with BGP
- Private: [Legacy 16-bit 64'512-65'535] [32-bit 4'200'000'000-4'294'967'294]

PEERING / NEIGHBORS

- Two routers with a BGP TCP session (port 179) are called peers/neighbors
- Each BGP router is a **BGP speaker**
- Supports CIDR, route aggregation; decisions based on policies/rules

PATH ATTRIBUTES

Used for route control and policy enforcement
Well-known mandatory: Always present (e.g., AS-Path, Origin, Next Hop)
Well-known discretionary: Optional but recognized by all (e.g., Local Pref)
Optional transitive: Passed between ASes (e.g., Community, Aggregator)
Optional non-transitive: Not passed across ASes (e.g., MED, Weight)
NLRI: Routing table info: prefix, prefix length, and associated path attributes

MESSAGES

OPEN: Establishes session; includes version, ASN, Hold Time, BGP Identifier, optional params. **Hold Time:** Heartbeat in seconds (Cisco def=180s), reset by KEEPALIVE/UPDATE; 0 = session down. **BGP Identifier:** 32-bit Router-ID, manually set or highest loopback/active IP; used for loop prevention
KEEPALIVE: Sent every 1/3 of Hold Time; ensures neighbor liveness
UPDATE: Advertises new routes, withdraws old ones, or both; includes NLRI (prefix + path attributes); can act as KEEPALIVE
NOTIFICATION: Sent on session error (e.g. hold timer expired); terminates session immediately

NETWORK STATEMENTS

Advertise a specific network prefix to BGP peers (only the best is advertised, even if multiple exist). Must exist in the RIB so that it's imported in BGP table
Connected: next-hop: `0.0.0.0`, origin attribute: `1`, weight: 32'768 (Cisco)
Static/RP: next-hop: NH in RIB, MED: IGP metric, rest same as above

BEST PATH CALCULATION

- BGP maintains all received paths per prefix but advertises only the best one
- Best path is installed in RIB; recalculated on:
  - Next-hop reachability change
  - Interface failure to eBGP peer
  - Redistribution change
  - New/withdrawn path received
- **Influence:**
  - Outbound BGP policy → inbound traffic behavior
  - Inbound BGP policy → outbound traffic behavior

BEST PATH SELECTION (IN ORDER):

1) Prefer highest **Weight (Cisco-specific**, local to router)
2) Prefer highest **Local Preference** (global within AS)
3) Prefer routes **originated by the router** (only small i in path, NH `0.0.0.0`)
4) Prefer **shorter AS path** (only length is compared)
5) Prefer **lowest origin type:** IGP < EGP < Incomplete (I on origin)
6) Prefer **lowest MED** (Multi-Exit Discriminator) (also called metric)
7) Prefer **external (EBGP)** over **internal (IBGP)**
8) For iBGP: prefer path with **lowest IGP metric** to next-hop
9) For eBGP: Prefer **oldest** (more stable) path
10) Prefer **lowest BGP router ID**
11) Prefer path from **lowest neighbor IP address**

ROUTE FILTERING

- Filters control which routes are received/advertised
- Used for security, traffic shaping, memory optimization
- Tools: **prefix-list** (IP), **filter-list** (AS-path), **route-map** (flexible match/set)

COMMUNITIES

32b optional, transitive tag (e.g. `ASN:value`, `65000:100`). Used to mark routes for policy control across ASes. Can be added, modified, or removed at each hop

ENTERPRISE CONNECTIVITY OPTIONS

Single/Dual: denotes how many **links** there are
Multi-Homed/-Homed: denotes how many **ISPs** are connected
Single-Homed: One ISP, one link (BGP or static). Simple but no redundancy
Dual-Homed: One ISP, two links (/routers). Redundancy within same provider
Multihomed: Multiple ISPs. improved redundancy and routing control, but avoid being a transit – advertise only customer-owned prefixes
Dual-Multihomed: Multihomed, but two links per ISP. Most redundancy

TRAFFIC ENGINEERING (TE) - INBOUND VS OUTBOUND

Local Pref: [OUT] [highest ↑] Lower local pref on backup path
MED: [IN] [lowest ↓] Higher MED on backup path
AS-Path Prepending: [IN] [shortest ←] Add multiple own ASN on backup path
TE Limitation: AS controls [OUT], [IN] control limited - ISPs may ignore
TE with Aggregate: Prefer primary ISP with summarized routes; advertise specific prefixes on backup for failover
Transit: One ISP provides reachability to all destinations
Peering: ISPs provide each other reachability to portions of their routing table

TRANSIT VS. PEERING

Transit: ISP provides full reachability (paid relationship)
Peering: ISPs exchange selected routes; equal relationship, usually unpaid
AS Path Filtering: to avoid getting transit: ip as-path access-list 10 permit ^$

INTERNET EXCHANGE POINT (IXP)

Facility where networks exchange traffic via BGP peering, ***routing policies*** determine which routes are advertised/accepted. Reduces transit costs, latency, and offloads upstream links

PUBLIC PEERING

Members peer via shared switch fabric and a ***route server***, which distributes routes but stays out of data path (NEXT_HOP unchanged). Simplified setup (one legal contract), minimal policy control; one BGP session to route server

PRIVATE PEERING

Direct BGP sessions between two parties ***(1:1)***, requires one session and legal contract per peer. May use **public or private** interconnects. Full policy control per neighbor. **Examples:** Equinix, SwissIX (non-profit)

RESOURCE PUBLIC KEY INFRASTRUCTURE (RPKI)

- Framework to validate that prefix is legitimately originated by a specific ASN
- Prevents route **hijacking** and accidental mis-originations (route **leaks**)
- RPKI validates origin only – not the AS-path

KEY COMPONENTS

Trust Anchors (TAs): Root CAs of the 5 RIRs; issue certs for resource holders
ROA – Route Origin Authorization: Digitally signed object, authorizes ASN to announce prefix (AS, prefix that AS can originate, max prefix length)
RPKI Validators: downloads+verifies+stores Validated ROA Payloads (VRPs)

RPKI VALIDATION STATES

Valid: Prefix + ASN match ROA
Invalid: Prefix found, but ASN mismatch or prefix too long
Not Found / Unknown: No matching ROA

VALIDATOR DETAILS

- Use TALs (Trust Anchor Locators) to fetch data from RIRs
- Validate cryptographic signatures (via X.509 certs with RFC 3779)
- Outputs VRPs; invalid objects are discarded
- Update frequency: >=24h (recommended), ~30–60min (practice)
- RRDP (RFC 8182) replacing rsync (uses HTTPS)

MONITORING

Event tracking, BGP hijack detection, Route leak detection, RPKI status check, Reachability tracking, AS path change tracking, AS path visualization

MULTICAST

Use-cases 1-to-Many: Streaming, software updates, music-on-hold
Use-cases Many-to-Many: Gaming, VR, stock data, group chat
Benefits: Efficient bandwidth, lower server/CPU load, no redundancy
Properties: UDP-based. Apps must handle drops, duplicates, out-of-order
Source: Sends to group IP; doesn't need to join
Receiver: Must explicitly join group to receive traffic

MULTICAST ADDRESS RANGES

`224.0.0.0 – 224.0.0.255`: Link-local, TTL=1 (not forwarded by routers) (IGMP)
`224.0.1.0 – 224.0.1.255`: Reserved by IANA, routable (NTP)
`232.0.0.0 – 232.255.255.255`: Source-Specific Multicast (SSM)
`235.0.0.0 – 238.255.255.255`: Globally scoped multicast (routable)
`239.0.0.0 – 239.255.255.255`: Administratively scoped (private space)

BROADCAST BASICS

L3 routes between subnets; L2 floods within same subnet
L2 (Bridging): MAC `ffff.ffff.ffff` switches flood to all ports in VLAN
L3 (Routing): `255.255.255.255`: local broadcast, not routed. Directed broadcast for specific subnet (e.g. `10.1.1.255`): can be routed if enabled

L2 MULTICAST: MAC MAPPING

1) **Get IP:** Example multicast IP address: `239.5.5.5`
2) **Convert to Binary:** `239.5.5.5` = 11101111.00000101.00000101.00000101
   → Take only the last 23 bits: 00000101.00000101.00000101
3) **Map to MAC Prefix:** Fixed 0100.5E → Map last 23 bits to: 0100.5E05.0505

INTERNET GROUP MANAGEMENT PROTOCOL (IGMP)

Purpose: Manages group membership for IPv4 multicast on each segment
IGMPv1: Basic join via query-response mechanism, no explicit group leave. Router sends ***membership queries*** every 60s to `224.0.0.1`, removes group after timeout if no report received. Receiver doesn't know multicast source
IGMPv2: Adds ***Leave Group*** message. Supports ***Group-Specific Queries*** e.g. when someone leaves group ('anyone still interested in group xy?'), reducing broadcast overhead. Receiver still doesn't know who the source is
IGMPv3: Adds ***source filtering*** (Include/Exclude lists). Enables ***Source-Specific Multicast (SSM)*** – receiver requests traffic only from selected source(s), **no need for Rendezvous Point** (RP) anymore. Adds support for application-level access control and filtering. Can also be used in ASM (Any Source Multicast), but mainly with SSM. (`224.0.0.22` all IGMPv3 routers)

IGMP SNOOPING

Without snooping: multicast = broadcast on VLAN
With snooping: switch listens to IGMP messages & builds a forwarding table
Default: snooping is enabled; switch needs IGMP Query to operate

REVERSE PATH FORWARDING (RPF)

To avoid loops, verifies that a multicast packet arrives on the interface that a unicast packet destined for the multicast source would be forwarded out of.

RENDEZVOUS POINT (RP)

Used in Shared Tree (`*,G`) setups with PIM-SM, acts as common meeting point. **RPF check** is performed toward the **RP** (not the source). Once the **source is known**, routers may **switch** to a Source Tree (`S,G`).

SHARED TREE VS. SOURCE-BASED TREE

Source-Based Tree (`S,G`): Built and added to ***route table*** when router receives an (`S,G`) join/report from IGMP host. `S` = known source; `G` = group
Shared Tree (`*,G`): IGMP host sends membership report (IGMP Join), Router adds (`*,G`) entry to ***multicast routing table*** `*` = 'any source'/unknown

PROTOCOL INDEPENDENT MULTICAST (PIM)

Only between routers. Relies on unicast routing table (RIB) for multicast forwarding decisions ⇒ protocol-independent. Uses RPF.

PIM DENSE MODE (PIM-DM) – NO RP

Push Model: Floods multicast to all interfaces; prunes where no receivers
(1) Flooding: Source sends traffic to multicast-enabled links using unicast RIB
(2) Distribution Tree: Initially entire network (shared tree rooted at source)
(3) Prune Messages: Routers without interested receivers send prunes upstream to remove themselves from the tree
(4) State Maintenance: Routers track source, receivers, interfaces to forward/prune per group

PIM SPARSE MODE (PIM-SM) – ASM, SSM

Pull Model: Multicast traffic is only sent where requested. Works with IGMP to detect interested receivers, uses unicast routing for forwarding
(1) Join/Prune: Routers send explicit join/prune messages to request or stop receiving multicast for group (G) to other routers
(2) Forwarding: Routers only forward multicast packets for group (G) on interfaces from which explicit joins were received

ANY SOURCE MULTICAST (ASM)

- Works with **IGMPv1** or **IGMPv2** → receiver does not know the source
- Receiver sends ***IGMP Join*** (`*,G`) to its first-hop router
- First-hop router forwards ***PIM Join*** (`*,G`) hop-by-hop toward the Rendezvous Point (RP)
- RP acts as a common meeting point for sources and receivers
- Sources send multicast traffic to the RP via a ***PIM Register*** tunnel
- All routers in the multicast domain must know the RP location

SOURCE SPECIFIC MULTICAST (SSM)

- Receiver subscribes using **IGMPv3**, providing both source (`S`) and group (`G`) to the first-hop router
- No Rendezvous Point (RP) required → **PIM-SSM builds only** (`S,G`) Shortest Path Trees (SPT)
- **No shared tree** (`*,G`) model used; SSM is source-directed
- IANA reserved ***232.0.0.0/8*** for SSM in IPv4
- Join messages are forwarded hop-by-hop toward the source to establish forwarding path
- Uses unicast routing table (RPF) to maintain loop-free delivery

PIM SPARSE-DENSE MODE

PIM Sparse Mode: Pull model – multicast traffic is forwarded only on request
PIM Dense Mode: Push model – traffic is flooded everywhere, then pruned
Sparse-Dense Mode: Supports both modes per multicast group; choice depends on RP availability

PIM-DM VS. PIM-SM

PIM is protocol-independent: it relies on the unicast routing table for RPF checks and to forward joins toward the source or Rendezvous Point (RP).

- **PIM Dense Mode (DM)**
  - "Push" model
  - Floods multicast traffic throughout the network
  - Prunes back where traffic unwanted
- **PIM Sparse Mode (SM)**
  - "Pull" model
  - Traffic sent only on request
  - Requires explicit ***Join*** messages

Usage Recommendation:
Dense Mode: Best for small or tightly scoped networks where most devices need multicast
Sparse Mode: Preferred for large-scale or distributed environments where multicast receivers are few or spread out

NETWORK DESIGN

Pillars: Scalability, Speed, Availability, Security, Manageability → overall cost

AVAILABILITY

MTBF: Mean Time Between Failures, **MTTR:** Mean Time to Repair
MTBF combined: $\left(\sum_{n=1} {}^{1}/_{\mathrm{MTBF}_n}\right)^{-1}$ **MTBF parallel:** $\sum_{n=1} {}^{\mathrm{MTBF}_n}/_{n}$
Availability: ${}^{\mathrm{MTBF}}/_{\mathrm{MTBF}+\mathrm{MTTR}}$ **More availability:** $\mathrm{MTTR} \to 0 \wedge \mathrm{MTBF} \to \infty$

REDUNDANCY

- Adds reliability, decreases MTBF but increases MTTR and complexity
- Balance: resilience vs. manageability

Backup Paths: Duplicate devices/links on primary path, build extra links for redundancy, consider backup link capacity, consider failover speed
Load Balancing: ECMP, EtherChannel, Port-Channel

HIERARCHICAL DESIGN

Access: Connect end devices, high port count, port security, L2, QoS marking
Distribution: L3, policy control, HSRP/VRRP, loop protection, small fault domain
Core: High-speed backbone, L3 only, no policy, scalable/redundant, no security
Collapsed Core: Combines Core + Distribution (small/medium networks)

FLAT DESIGN

- Any-to-any; small networks, MPLS/LAN setups
- Lacks scalability and control

FABRIC DESIGN

Modern design using **Underlay/Overlay**
Underlay: transport (e.g., IP, MPLS)
Overlay: logical virtual topology (e.g., EVPN)

ENTERPRISE CAMPUS

- 100s/1000s of users, multiple buildings, one physical location
- Multiple interconnected LANs, connected via Ethernet and Wireless

FHRP – FIRST HOP REDUNDANCY PROTOCOLS

- Ensures default gateway is always reachable, PCs can only have one
- Enables fast failover during router failure

VRRP – Virtual Router Redundancy Protocol (Multivendor): Shared virtual IP, real MACs per router, Master router handles forwarding
HSRP – Hot Standby Router Protocol (Cisco): Shared virtual IP + virtual MAC, One active, others in standby → faster
GLBP – Gateway Load Balancing Protocol (Cisco): Load balancing + redundancy, Shared virtual IP + multiple virtual MACs, Roles: **AVG:** Answers ARP and sends virtual MAC addresses of AVFs **AVF:** Forwards traffic

DATA CENTER DESIGNS

ToR - Top of Rack: switches per rack, less cabling, easy expansions/exchanges per 'rack', scalable glass fiber, ideal for high service density (full racks). But more switches, more ports, more L2 Srv-2-Srv traffic, more STP to be managed
EoR - End of Row: 1 switch per row, less switches, higher utilization of ports, switches all at one place, better L2 availability between racks. But more cabling

TRAFFIC DIRECTIONS

North–South: Between external networks (client-server, in/out DC)
East–West: Within DC (e.g., server-server, storage)

THREE-TIER DC

- Access – Aggregation – Core (like Hierarchical)
- Optimized for North–South traffic
- Not ideal for East–West communication

LEAF-SPINE ARCHITECTURE

- Two-tier: Leaf switches (access) connect to Spines (core)
- High performance, low latency
- Scalable, ideal for East–West traffic

VXLAN

Issues of L2: STP, Max amount of VLANs (4094), Large MAC Address tables
VXLAN (Virtual Extensible LAN): Tunnels Ethernet (Layer 2) over IP using MAC-in-UDP encapsulation (Port 4789). For flexible and scalable network segmentation.
VNID (VXLAN Network Identifier): 24-bit identifier (up to 16 million segments) that defines the VXLAN broadcast domain.
VTEP (Virtual Tunnel Endpoint): Device (switch, router, or host) responsible for encapsulating/de-encapsulating VXLAN traffic.
NVE (Network Virtual Interface): Logical interface on a VTEP used for VXLAN tunnel operations.

TUNNEL

- VXLAN establishes IP tunnels between VTEPs to extend Layer 2 networks across Layer 3 boundaries.
- VXLAN enables both L2 and L3 VPN functionality in overlay networks.
- VXLAN traffic is encapsulated in UDP (default port: ***4789***).

FRAME FORMAT

- Ethernet frame ⟶ VXLAN Header ⟶ UDP ⟶ Outer IP Header.
- The VXLAN header contains the 24-bit VNID and flags.
- Outer headers allow Layer 2 frames to traverse IP underlay networks.

VIRTUAL NETWORK IDENTIFIER (VNI)

- 24-bit VXLAN Network Identifier uniquely defines VXLAN segments.
- Replaces traditional VLAN IDs (12-bit), enabling ~16 million logical segments.
- Used by VTEPs to map traffic into corresponding Layer 2 domains.

VXLAN TUNNEL ENDPOINT (VTEP)

- Connects the overlay (VXLAN) and underlay (IP) networks.
- **Types:**
  - **Software VTEP:** Located on hypervisors using virtual switches.
  - **Hardware VTEP:** Located on routers/switches with ASICs for performance.
- **Interfaces:**
  - **VTEP IP Interface:** Connects to the underlay network and handles encapsulation.
  - **VNI Interface:** Virtual interface per segment (like SVI); handles segregation of Layer 2 domains.

MAC ADDRESS LEARNING

On control plane: happens proactively
On data plane: ad-hoc with flooding

- Each VTEP maintains a VXLAN mapping table linking destination MAC addresses to remote VTEP IPs.
- **Learning via ARP:**
  - Host H1 sends ARP request, switches learn H1's MAC.

- ARP request is flooded to H2.
- H2 responds; switches learn H2's MAC.
- **Learning Methods:**
  - **Static VXLAN:** Manual MAC-to-VTEP mappings. Doesn't scale well; BUM traffic is inefficient.
  - **Multicast VXLAN:** VTEPs join multicast groups per VNI. Scales better, offloads BUM replication. 20+ VTEPs = there is too much traffic, doesn't scale well
  - **MP-BGP EVPN:** Modern solution using BGP as control plane. Dynamically learns MAC/IP info.

EVPN

Overcome flood-and-learn limitations, doesn't rely on data plane learning, utilizes robust control plane MP-BGP, works with different encapsulation techniques (VXLAN, MPLS), excellent scalability, l2 and l3 Support.

MP-BGP EVPN (MULTIPROTOCOL BGP FOR ETHERNET VPN)

- Enables protocol-based VTEP discovery and host reachability via control-plane learning
- Reduces flooding by replacing data-plane learning
- Extends BGP with multiprotocol capabilities (AFI/SAFI)
- Uses ***MP_REACH_NLRI*** and ***MP_UNREACH_NLRI*** for route advertisement and withdrawal

EVPN ROUTE TYPES

Type 2 – Host Advertisement: Advertises host MAC (mandatory), optionally IP, along with L2VNI and optionally L3VNI. Used for MAC learning, ARP suppression, and host mobility. Sent when host connects to VTEP.
Type 5 – Subnet Advertisement: Advertises IP prefix + prefix length with L3VNI. Used for inter-subnet routing. VTEP redistributes connected/static/dynamic IP routes. Additional attributes: L3VNI, extended communities.

HOST DELETION & MOVE

Host Deletion: When a host detaches, its ARP (default: 1500s) and MAC entry (default: 1800s) time out on the VTEP. Upon aging, the VTEP withdraws the host's MAC/L2VNI and IP/L3VNI advertisements.
Host Move: When a host moves to a new VTEP, the new VTEP advertises updated reachability with a higher move sequence number. The old VTEP withdraws its entry, completing the migration.

ROUTE DISTINGUISHER (RD) VS. ROUTE TARGET (RT)

- **Route Distinguisher (RD):** Uniquely identifies VPN routes — allows same IP prefix to be used in different VPNs. Can be IPv4 or ASN ***Used to make routes unique in BGP (VPNv4/v6).*** Forms VPNv4 NLRI: ***RD:IPv4 prefix***
- **Route Target (RT):** Controls route import/export between VRFs. ***Used as extended BGP community.***
- **How RTs Work:**
  - A route is tagged with an RT when advertised by BGP.
  - Other VRFs import the route if the RT matches their import policy.
  - Allows overlapping or shared connectivity between tenants (e.g., shared services).
- **Format:** Typically in the form ***ASN:nn*** or ***IP:nn***, e.g., `65000:100`, `1:10`
- **Multiple RTs can be used:** A route can have multiple RTs for flexible policies (e.g., one RT for VPN, another for shared services)

EVPN (ETHERNET VPN) – L2

KEY FEATURES

- L2 bridging across L3 networks

BGP Control Plane: Distributes MAC info (no flooding)
VXLAN Overlay: Encapsulates L2 in L3 UDP (data plane)
Multi-Tenancy: via VNI segmentation
Redundancy: All-active multihoming, ECMP, fast convergence

USE CASES

- Multi-tenant datacenter interconnects (DCI)
- Extending L2 over WAN between remote sites
- Scalable, segmented L2 fabrics

BGP CONTROL PLANE

- PEs learn MACs from local CEs (data plane)
- MACs advertised via BGP (control plane)
- Uses Route Distinguishers and MPLS labels
- Remote PEs update L2 RIB/FIB with MAC and next-hop info
- Enables seamless L2 across IP/MPLS backbone

EVPN NLRI

- EVPN uses MP-BGP with specific AFI/SAFI
- Supports multiple route types and attributes
- Unsupported routes are dropped by BGP

AUTODISCOVERY VIA ROUTE REFLECTORS

- Route Reflector (RR) avoids full-mesh iBGP
- RR reflects EVPN routes to other PEs
- RR doesn't participate in EVPN or pseudowires
- RR needs only ***address-family l2vpn evpn***
- L2VPN RIB stores endpoint/VFI info for control plane
- ***BGP_UPDATE*** from spines contain ***ORIGINATOR_ID*** (origin leaf)

HOST DETECTION

- Host connects to VTEP → MAC learned locally
- VTEP advertises MAC + L2VNI via BGP EVPN
- MAC learning follows normal Ethernet semantics

INGRESS REPLICATION (IR)

- BUM traffic, when Multicast underlay network is not used, handle multi-destination traffic (ARP → unicast)

EARLY ARP TERMINATION (ARP SUPPRESSION)

- Avoids flooding ARP requests
- VTEP queries control plane for MAC/IP/VNI mapping
- If known → direct unicast (no broadcast)

SILENT HOST FLOW (FALLBACK)

- If IP/MAC unknown → ARP sent via **ingress replication**
- Replicated ARP request goes to remote VTEPs
- Only correct host responds → update reflected to all VTEPs
- Future traffic uses updated BGP mapping

VRF – VIRTUAL ROUTING AND FORWARDING

- Multiple isolated routing tables on one device
- Each tenant = one VRF → traffic isolation
- Supports independent policies per tenant
- Key for scaling and multi-customer separation

IRB – INTEGRATED ROUTING AND BRIDGING

- Enables inter-VLAN routing inside EVPN
- Avoids central gateway → no 'traffic tromboning'
- Two modes: Symmetric and Asymmetric

SYMMETRIC IRB (L2 + L3)

- Routing/bridging on ingress + egress VTEPs
- Uses L3 Transit VNI (same in both directions), One L3 VNI per VRF (Tenant)
- Scales well; clean separation of MAC and IP

ASYMMETRIC IRB (L2)

- Routing only on ingress, bridging on egress
- VXLAN uses destination VNI in both directions
- One L2 VNI per VLAN/Subnet
- Simple config, but requires all VLANs/VNIs on all VTEPs

DISTRIBUTED ANYCAST GATEWAY (DAG)

- Same gateway IP+MAC on all VTEPs
- Enables local default gateway for hosts
- Supports mobility + optimal forwarding

L3 HOST DETECTION

- Host sends ARP/ND to local VTEP
- VTEP learns MAC/L2VNI and IP/L3VNI
- Info is advertised in EVPN (control plane)

MPLS

- Label Switched Path (LSP) → pre-determined path across MPLS network
- advantage eBGP between PE-CE: No mutual redistribution, same routing process
- encrypt traffic flowing over MPLS L3VPN backbone? yes (e.g. bank)
- Unicast Reverse Path Forwarding (uRPF): checks source of each packet & verifies that source is in routing table
- control plane (e.g. OSPF) → to learn labels
- iBGP used to exchange NLRI (RD, RT, IPv4 Prefix, NextHop &VPN Label) between PE
- imp-null = networks are directly connected, no more label switching

WAN

Connects remote LANs via SPs for data/voice/video; key needs: bandwidth, control, design, resilience, mgmt. **Requirements:**
Bandwidth: App needs, peak usage, reserve for VoIP
Control/Security: Trust provider? No full control
Availability: Redundancy, SLA for failures
Mgmt: Inband vs out-of-band

PRIVATE WAN

Point-to-Point: Leased L2 line (Ethernet); monthly fee; private circuit
Dark Fiber: Physical fiber lease; costly; ISPs prefer selling lambdas
Connection-oriented: Predefined path, packets carry IDs (ATM, Frame Relay)
Connectionless: No setup; full address in each packet (Ethernet, MPLS VPN)

TERMINOLOGY

CE - Customer Edge: no knowledge of MPLS, no labels; connected to PE
PE - Provider Edge: connected to CE; runs iBGP and LDP; uses VRFs
P - Provider or LSR(Label Switch Router): inside MPLS VPN, no CE connection; forwards labels

DATABASES, PLANES

RIB (Routing IB (Information Base)): Learned prefixes from routing protocols
FIB (Forwarding IB): Built from RIB; only best routes for forwarding
LIB (Label IB): All label mappings; 1 label per prefix
LFIB (Label Forwarding IB): Built from LIB; used for actual forwarding decisions. (L)FIB only contains currently best LSP (decision: Routing Protocol)
Control Plane: Builds routing/label tables (RIB, LIB)
Data Plane: Forwards packets (FIB, LFIB); pushes/swaps/pops labels (see below)

MPLS HEADER

4-byte header before IP:
- **Label (20b)** – actual MPLS label
- **EXP (3b)** – QoS/CoS, Now called Traffic Class (TC)
- **S-bit (1b)** – bottom of label stack indicator, 1 = True = last label before IP header
- **TTL (8b)** – time-to-live (eq toual IP TTL)

TTL MPLS

- ingress PE router decrements IP TTL field & copies packet's IP TTL field into new MPLS TTL
- P routers decrements MPLS TTL
- egress PE router decrements MPLS TTL, pops final MPLS header, copies IP TTL
- traceroute receive ICMP Time Exceeded, Provider doesn't want to expose MPLS network to fix: disable MPLS TTL propagation (on PE), PE set MPLS TTL = 255, egress leaves PE original IP TTL unchanged
- = MPLS network appears as single router hop from TTL perspective

LABEL DISTRIBUTION PROTOCOL (LDP) - CONTROL PLANE

- Distributes labels to neighbors using control plane
- **Hello messages:** Sent via UDP (Port 646) to `224.0.0.2` to discover neighbors
- TCP (Port 646) connection is used to exchange label bindings (prefix to local label)
- Routers advertise all local bindings after TCP session is up
- Label mapping used to build LIB → LFIB
- LDP router ID must be reachable (via routing table)
- Each router manages local labels independently

MPLS L3 DATA PLANE

- VPN traffic uses **2 labels** (stacked):

Outer label: Transport label (LDP); identifies LSP between ingress/egress PE
Inner label: VPN label (MP-BGP); identifies customer VRF
Push: Ingress PE; classify and label packets
Swap: P router; replaces label, forwards based on new label
Pop: Egress PE removes label; sends original packet to CE
Penultimate Hop Popping: MPLS feature, penultimate router removes the outer MPLS label before forwarding to egress PE, default enabled, ISPs disable it

VRF TABLES

- VRF = Virtual Routing and Forwarding table (Virtual router inside a PE. Maintains isolated RIB + FIB per customer.)
- Stores separate routing info per customer (VPN isolation)
- Exists per MPLS-aware PE router; one per attached customer
- Contains: RIB, FIB, and separate routing process per CE

VPNV4

- 64-bit RD + 32-bit IPv4 = 96-bit VPNv4 prefix
- transferring VPNv4 between PE router → Multiprotocol iBGP (MP-iBGP)

OVERLAY TECHNOLOGIES

MODERN PROVIDER NETWORK

MPLS: Label-based forwarding (fast, scalable)
LDP: Distributes labels for MPLS paths
IGP: Underlay routing (e.g., OSPF, IS-IS)
MP-BGP: Extends BGP to carry VPNv4/v6, EVPN routes

DRAWBACKS OF TRADITIONAL NETWORKS

Control Plane: LDP/RSVP-TE adds complexity
Scalability: Per-flow/path state limits growth; LSP and signaling overhead increase rapidly
OAM:
- **Troubleshooting:** Traceroute less useful in MPLS; labels hide topology
- **Traffic Eng.:** LDP lacks TE; relies only on IGP cost

Fast Reroute: Limited coverage; microloops possible

SEGMENT ROUTING (SR)

Source routing: Sender defines full path using Segment List (Segment = Instruction)
SID = Segment Identifier: Each SID = 1 instruction (e.g., forward via ECMP, specific iface, or to a service)
State in packet: No per-flow state in network; intermediate nodes follow SID instructions
No new control plane: Uses existing protocols (OSPF, IS-IS, BGP) width extensions; no LDP or RSVP-TE needed
Segment List: Ordered SID list carried in packet header; defines full route
Simple but powerful: Enables TE, fast reroute, policy routing

SEGMENT LIST OPERATIONS

Push: Insert SIDs into packet; set active SID (top of list)
Continue: Active SID not yet completed; keep processing it
Next: Current SID completed; activate next SID in list

SEGMENT SIGNIFICANCE

Global Segments: Known and supported by all SR nodes in the domain, Installed in forwarding tables across the network (e.g. "Forward packet according to shortest path to Node1")
Local Segments: Defined and installed only on originating node, Not forwarded by others, but must be understood network-wide (e.g. "Forward packet on interface to Node2")
Global segments: Defined in the SR Global Block (SRGB) and should be consistent across all nodes; **local segments** are defined in the SR Local Block (SRLB) and are specific to the local SR node.

SR CONTROL PLANE SEGMENT TYPES

IGP Prefix Segment: Global SID tied to IGP prefix (multi-hop); all nodes install forwarding entries
IGP Node Segment: Global SID for a specific node (shortest-path forwarding)
IGP Anycast Segment: Global SID for a group of nodes; traffic sent to nearest
IGP Adjacency Segment: Local SID; direct link to neighbor
L2 Adjacency SID: Local SID for Layer-2 segment (e.g., Ethernet link)
Combining Segments:
- End-to-end paths can mix IGP and BGP segments
- Traffic to BGP Anycast → more ECMP in data centers

SR-MPLS

- Reuses existing MPLS data plane — no hardware change needed
- Segments = MPLS labels; Segment List = label stack (top = active)
- Segments distributed via IGP/BGP; no LDP required (interoperable if needed)
- Supports both IPv4 and IPv6 networks

BENEFITS OF SEGMENT ROUTING

Benefits: Simplification (removes protocols, simple operations, admin and mgmt), enhanced Traffig eng. (Delay, Bandwidth, Packet Loss, TE metric, Controller, Source-Node), Seamless deployment, Robust, Network Innovation (zB Container Networking)
Source Routing: Balances distributed intelligence with centralized optimization
TI-LFA: Fast reroute technique; protects against link/node failure with microloop avoidance and no pre-calculation dependency
Traffic Engineering (TE): Optimizes network performance by analyzing and controlling data flow to reduce congestion and improve QoS
Service Function Chaining (SFC): Chains SDN services in order; automates traffic between VNFs and optimizes routing for performance

QOS

Internet is best effort: no guarantees, no QoS; all traffic treated equally (net neutrality); simple, scalable, but no delivery/order assurance or prioritization

QOE & ROUTE PINNING

QoE – Quality of Experience: Perceived service quality from user perspective
Route Pinning: Keeps flow on a fixed path to prevent oscillation (don't switch immediately to "better" path)

NETWORK PERFORMANCE METRICS

Latency / Delay [ms]: Time for packets to travel src → dest (Voip < 150ms)
End-to-End Delay: Total time sender to receiver
One-Way Delay: From first bit sent to last bit received
Delay Components: Transmission delay (time to push onto link), Processing delay (lookup, queuing), Propagation delay (physical travel time)
Jitter [ms]: Variation in delay between packets, caused by re-routing/queuing (Voip<30ms), Calc: no queue - queued delay
Throughput: Rate of successfully delivered data
Packet Loss [%]: Dropped packets due to congestion or errors (Voip < 1%)
Bandwidth [Gbit/s]: Maximum transfer capacity of a link

QUEUING ALGORITHMS

FIFO (First-In First-Out): Basic, no prioritization
Priority Queuing (PQ): Multiple queues, serve highest first; others may starve
Round-Robin: One packet per queue in turn (fair, but ignores priority)
Weighted Fair Queuing (WFQ): Round-Robin with weights, e.g., 2 packets from Q1, 4 from Q2
Class-Based WFQ (CBWFQ): WFQ with user-defined classes, queue limits, max bandwidth guaranteed or max % of bandwidth (logical queues based on IP Precedence only)
Low Latency Queuing (LLQ): Adds strict priority queue (priority class) to CBWFQ for delay-sensitive traffic (e.g. voice) (based on IP Precedence, DSCP, src, port, protocol…)

QUEUE MANAGEMENT

Tail Drop: Drops packets when queue full; huge interruption of traffic → same as no connectivity
TCP Global Sync: Many TCP flows back off and restart simultaneously → link underutilization
TCP Starvation: TCP slows down after drops, UDP doesn't → queues filled with UDP, TCP squeezed out
RED: Random early drops before full queue to prevent global sync and TCP collapse. Dropped TCP segments cause TCP sessions to reduce their windows sizes
WRED: RED + DSCP/EXP-based drop logic, prioritizes higher-marked traffic
DSCP / EXP: DSCP (6-bit in IP header) marks packets for QoS; used in DiffServ for classifying traffic. EXP (3-bit in MPLS label) serves same purpose within MPLS networks; often mapped from DSCP.

POLICING VS. SHAPING

Policing (Inbound mostly): Drops packets that exceed configured rate limits
Shaping (Outbound): Buffers packets to smooth traffic bursts and conform to profile

QOS MODELS

Best Effort: No guarantees, all traffic treated equally (follows Internet neutrality)
Integrated Services (IntServ): End-to-end QoS, per-flow resource reservation, precise but not scalable (uses RSVP)
Differentiated Services (DiffServ): Class-based, scalable approach using marking (e.g.,), no hard guarantees

TRAFFIC MARKING

L3 Marking: ToS byte → DSCP (6 bits) + IP Precedence (3 bits)
L2 Marking: Dot1q header → 802.1p CoS bits

MODULAR QOS CLI (MQC)

Class Map: Define traffic classes (e.g., match voice or video)
Policy Map: Define actions for each class (e.g., limit, shape, priority)
Service Policy: Apply policies to interfaces or directions (in/out)

CDN

Origin Server: Central content source (original files), usually in a datacenter
Edge / CDN Server (POP - Point of Presence): Geographically distributed, caches content
DNS Infrastructure: Directs users to optimal edge server (e.g. via Geo-Routing)

KEY BENEFITS

Latency Reduction: Nearby edge servers reduce round-trip time
Availability: Failover and redundancy in case of node failure
Scalability: Handles traffic spikes via load balancing
Cost Optimization: Reduces backend and transit load on origin
DDoS Protection: Edge servers absorb attacks → not all traffic on one server
Global Load Reduction: Less long-distance traffic across the Internet

REQUEST ROUTING TECHNIQUES

- Decides which edge server should serve a client request
- Goal: Best performance (e.g. proximity, load, responsiveness)

DNS-BASED GEO-ROUTING

- Each edge has a unique IP
- DNS server picks closest/optimal edge server based on:
  - Resolver IP location (not user!)
  - GeoIP DBs (MaxMind, IP2Location), load, latency, business rules
- Limitation: DNS Resolver != user location → can cause wrong choice

EDNS(0) AND CLIENT SUBNET EXTENSION (ECS)

- Resolver includes part of client IP in DNS request (e.g. /24 subnet)
- Authoritative DNS makes better decision based on actual client region
- Improves accuracy without revealing full IP

ANYCAST WITH BGP

- Same IP (e.g. `7.7.7.7`) advertised from multiple locations
- BGP routing decides which path is "best" (AS-path, local pref, etc.)
- No DNS logic or per-client decision — pure BGP convergence

Pros: Fast failover, simple, no app logic needed
Cons: Less control, BGP != best latency, route flapping risk

HTTP CACHING & HEADERS

- Caching is controlled via HTTP headers between clients, proxies, and servers

Cache-Control: Main directive (***no-cache***, ***no-store***, ***max-age***, ***must-revalidate***, etc.)
Expires: Absolute expiration time (older method, replaced by ***Cache-Control***)
ETag: Validator tag (version/hash), used with ***If-None-Match***
Last-Modified: Timestamp used with ***If-Modified-Since*** for revalidation
Age: Time (in seconds) since response was fetched from origin
Validation: Client uses ***ETag*** or ***Last-Modified***; server returns 304 if unchanged

OTHER INFOS

AD: Inter-protocol choice (e.g., OSPF vs RIP) → lower wins.
Cost/Metric: Intra-protocol choice (e.g., OSPF path A vs B) → lower wins.
Routing Preference Order (across protocols):
- Most specific prefix
- Lowest Administrative Distance
- Static default route

Administrative Distances: (Smallest Administrative Distance wins)

Protocol	*Distance*
Connected	0
Static (Interface)	1
Static (Next Hop)	1
BGP External	20
EIGRP Internal	90
OSPF	110
ISIS	115
RIP v1/v2	120
EIGRP External	170
BGP Internal	200

EVPN BGP ROUTING TABLE INFOS

* = Would not be there if it was L2 VNI BGP Routing Table
Route Distinguisher: `172.16.255.101:32777`
Route Type: 2
MAC Address Length: 48
MAC Address: 5254.00f8.29a8
***IP Address Length**: 32
***IP Address**: `10.10.0.100`
L2 VNI: 30010
***L3 VNI**: 50000
Remote VTEP IP Address: `172.16.254.101`
L2 Route Target: 1:10
***L3 Route Target**: 65000:50000

```
leaf-03# show bgp l2vpn evpn 10.10.0.100
BGP routing table information for VRF default, address family L2VPN
EVPN
Route Distinguisher: 172.16.255.101:32777
BGP routing table entry for
[2]:[0]:[0]:[48]:[5254.00f8.29a8]:[32]:[10.10.0.100]/272, version 19897
Paths: (1 available, best #1)
Flags: (0x000202) (high32 00000000) on xmit-list, is not in l2rib/evpn,
is not in HW
Advertised path-id 1
Path type: internal, path is valid, is best path, no labeled nexthop
Imported to 2 destination(s)
AS-Path: NONE, path sourced internal to AS
172.16.254.101 (metric 81) from 172.16.255.1 (172.16.255.1)
Origin IGP, MED not set, localpref 100, weight 0
Received label 30010 50000
Extcommunity: RT:1:10 RT:65000:50000 ENCAP:8 Router MAC:5254.00ca.69ae
Originator: 172.16.255.101 Cluster list: 172.16.255.1
```

TODO:

```
\begin{center}
  *Route Type 2:*
  \includegraphics[width=\linewidth]{route-type-2}
  *Route Type 5:*
  \includegraphics[width=\linewidth]{route-type-5}
\end{center}
```

PRÜFUNG VORJAHR

NETWORK DESIGN

3-tier campus network: Default Gateway (D), QoS marking (A), STP Root Port (A), HSRP, VRRP or GLBP (D), "Simple" (C), OSPF Totally Stub Area (D), High availability (C)
Campus Design: used to reduce size of L2 domain: EVPN, MPLS

REST

MP_REACH_NLRI: Next hop, MAC Address

- VXLAN is a data plane technology which encapsulates Ethernet frames in UDP datagrams to tunnel layer 2 frames over a layer 3 network.
- The underlay network is unaware of VXLAN devices that connect to the physical switches are unaware of VXLAN.
- A route distinguisher is used to uniquely identify a route in combination with the destination prefix.

TODO'S

OSPF
- passive interfaces
- flooding
- O N1 / O N2 routes
- route summarization
- synchronizing the lsdb

IS-IS
- ES-IS, IS-IS: details
- adjacencies
- **route leaking**
- summarization

BGP
- comparison to IGP
- Multihop sessions
- NLRI
- Aggregate impact (TE)
- rework
  - Best path calculation

INTERNET
- Routing policy
- IXP

AVAILABILITY
- everything

MULTICAST
- Link-local multicast (48)
- IPv4 MAC Address mapping
- IGMP and the querier/Role in Dense Mode/Challenges

TODO:
notes 37+