Evolving Technologies

By Orhan Ergun

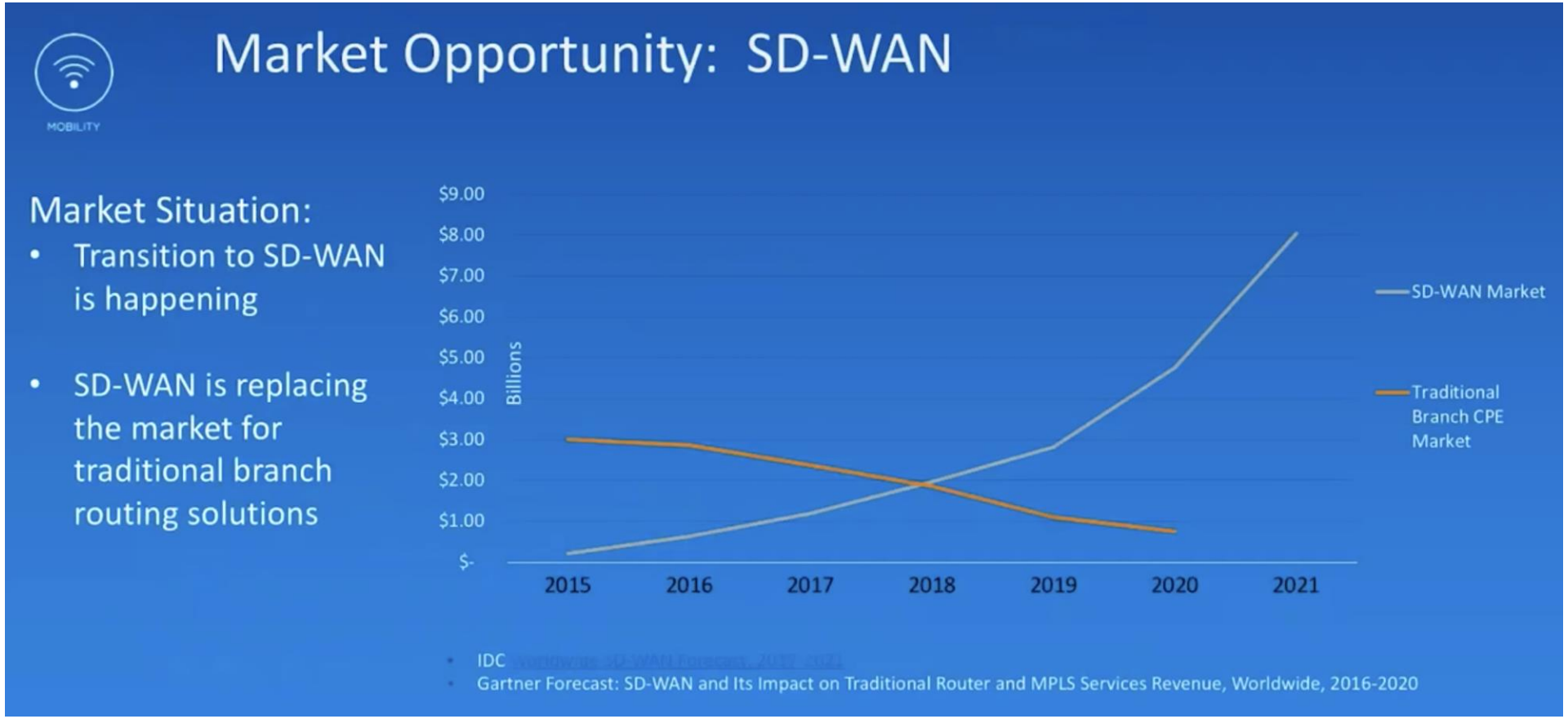
Evolving Technologies Course Outline

- SD-WAN
- Cloud Computing
- Edge and Fog Computing
- IPTV
- CDN – Content Delivery Networks
- Wireless Local Area Network Design
- IOT – Internet of Things
- NFV
- Intent Based Networking
- Artificial Intelligence
- Machine Learning and Deep Learning
- 5G
- Tactile Internet
- Segment Routing
- DEVOPS

SD-WAN Software Defined Wide Area Network – What is SD-WAN ?

- SD-WAN is an acronym for software-defined networking in a wide area network (WAN)
- SD-WAN simplifies the management and operation of a WAN by decoupling the networking hardware from its control mechanism
- This concept is similar to how software-defined networking implements virtualization technology to improve data center management and operation

SD-WAN Market and Estimated Market Size



SD-WAN – Why SD-WAN ?

- A key reason of SD-WAN is to allow companies to build higher-performance WANs using lower-cost and commercially available Internet access, enabling businesses to partially or wholly replace more expensive private WAN connection technologies such as MPLS

Lots of buzzwords! Which one is SD-WAN?

Application based routing

Multipath routing

Quality of Experience

Centralized orchestration

Visibility and Analytics

Automation

Encrypted tunneling

Access independent

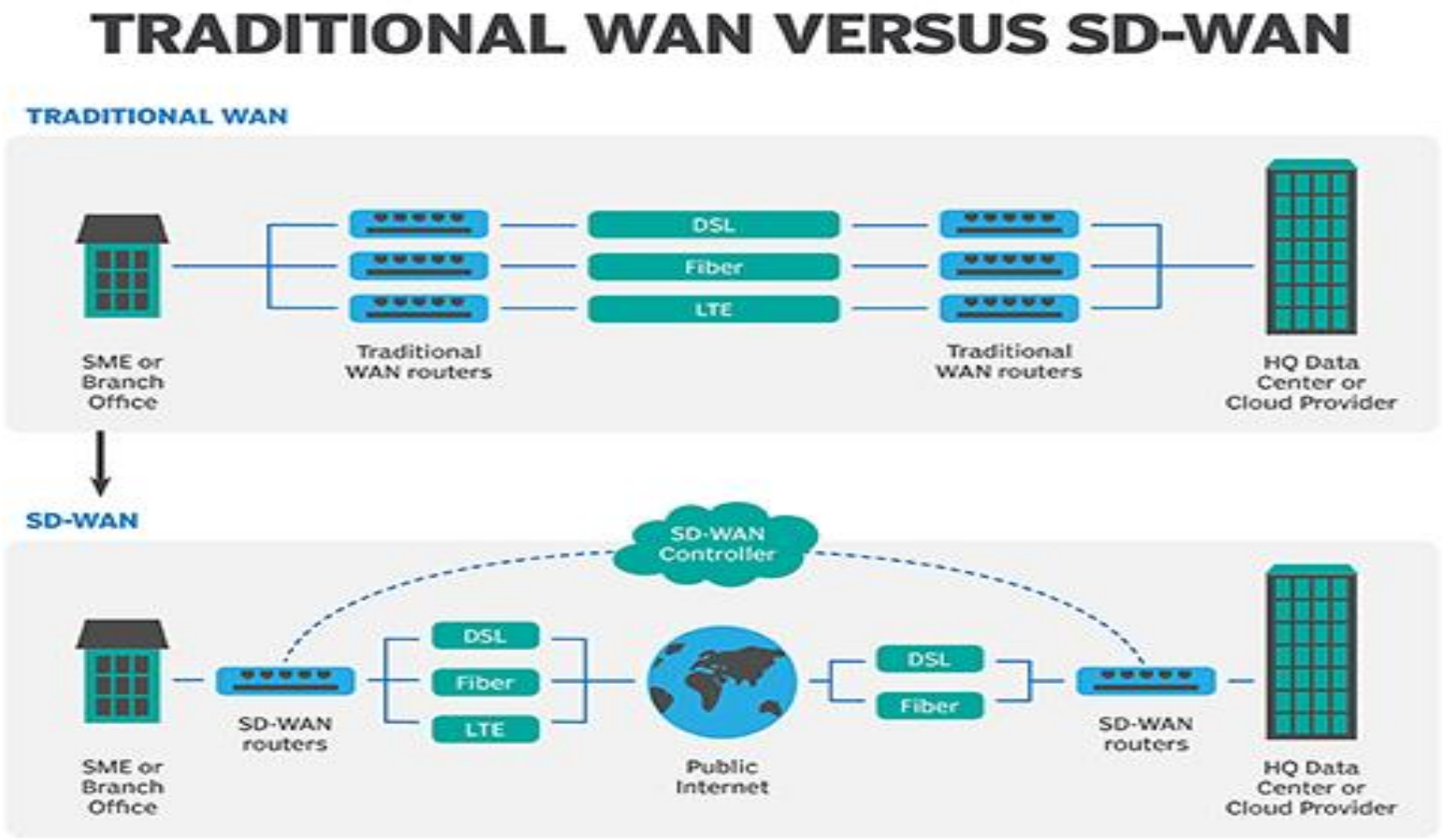
ZeroTouchDeployment

Security

Reduce # HW on site

Cloud integration

Traditional WAN vs. SD-WAN

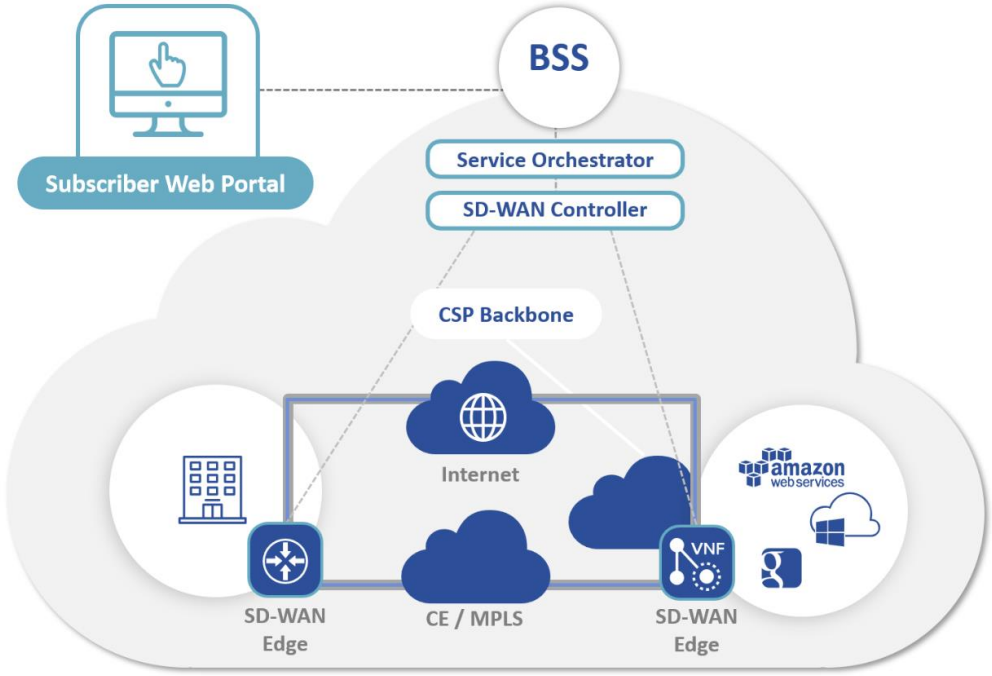


Why SD-WAN – MEF Definition

- MEF's SD-WAN service definition standard describes requirements for an application-aware, over-the-top WAN connectivity service that uses policies to determine how application flows are directed over multiple underlay networks regardless of the underlay technologies or service providers who deliver them

MEF SD-WAN Service Definition

SD-WAN Service Constructs



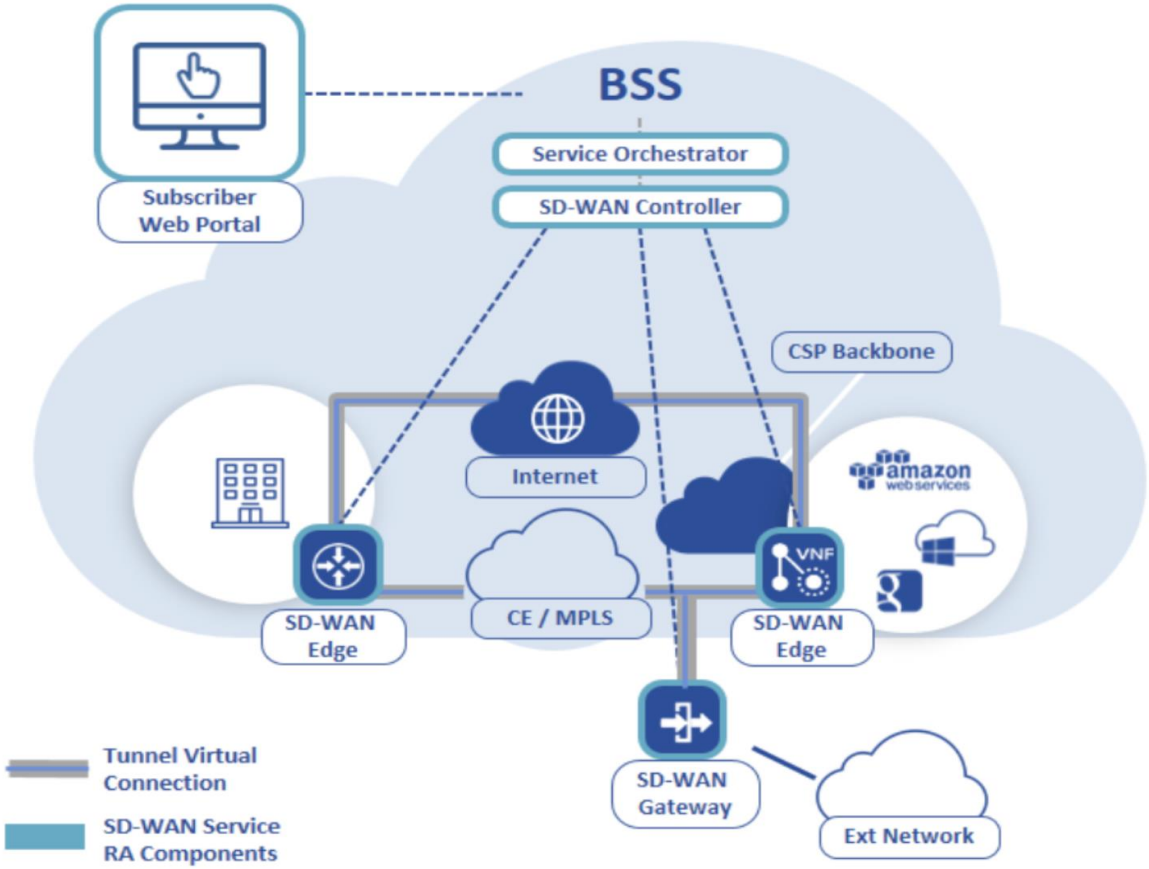
SD-WAN Edge
Physical or virtual

SD-WAN Controller
Centralized management of SD-WAN edges & gateways

Service Orchestrator
Lifecycle Service Orchestration of SD-WAN and other services

Subscriber Web Portal
Subscriber service ordering and modification

SD-WAN Components – Common Architectural Elements



SD-WAN Service Components. Source: MEF

SD-WAN 5 Common Elements:

Edge
Controller
Gateway
Orchestrator
Web Portal

SD-WAN Components – Common Architectural Elements- SD-WAN Edge

- SD-WAN Edge, which could either be formed by a physical device, such as a CPE, or by a virtual CPE that's based on a virtual network function (VNF)

- The SD-WAN Edge performs several critical functions

Source: cisco.com

SD-WAN Components – Common Architectural Elements- SD-WAN Edge

- It acts as the security-policy enforcer and conducts WAN optimization tasks that include data deduplication, compression, and packet buffering

- It also creates and removes encrypted tunnels over underlay networks, whether it's a wired or wireless connection

SD-WAN Components – Common Architectural Elements- SD-WAN Edge

- Since SD-WAN Edges often connect to public Internet WANs, they would include, at a minimum, some NAT and firewall capabilities

- Most of the time actual cost of the deployment comes from SD-WAN Edges as there are much more Edge device in the network than Centralized Core Devices (Gateways , Controllers , Orchestrator)

SD-WAN Controller

- SD-WAN Controller centralizes management to the SD-WAN Edge and to the SD-WAN Gateway
- The SD-WAN Controller provides physical or virtual device management for all SD-WAN Edges and SD- WAN Gateways associated with the controller
- This includes, but is not limited to, configuration and activation, IP address management, and pushing down policies onto SD-WAN Edges and SD-WAN Gateways

SD-WAN Controller

- The SD-WAN controller maintains connections to all SD-WAN Edges and SD-WAN Gateways to identify the operational state of SD- WAN tunnels across different WANs and retrieve QoS performance metrics for each SD-WAN tunnel

- These metrics are used by the Orchestrator

SD-WAN Orchestrator

- The Service Orchestrator provides the service management of the SD-WAN service lifecycle including service fulfillment, performance, control, assurance, usage, analytics, security and policy

- For example, the Service Orchestrator is responsible for configuring the end-to-end SD- WAN managed service between SD-WAN Edges and SD-WAN Gateways over one or more underlay WANs, e.g., Internet and MPLS, setting up application-based forwarding over WANs based on security, QoS or business or intent-based policies

SD-WAN Gateway

- The SD-WAN Gateway is a special case of an SD- WAN Edge that also enables sites interconnected via the SD-WAN to connect to other sites interconnected via alternative VPN technologies, e.g., CE or MPLS VPNs

SD-WAN Gateway

- There are two ways to deliver an SD-WAN service to sites connected via another VPN service.

- One way requires an SD- WAN Edge to be placed at each subscriber site connected to the VPN service so SD-WAN tunnels can be created over the VPN

SD-WAN Gateway

- Another way is to use an SD-WAN Gateway
- In this scenario, an SD-WAN Gateway initiates and terminates the SD-WAN tunnels like an SD-WAN Edge and initiates and terminates VPN connections to and from sites interconnected by the VPN
- This approach enables sites interconnected via SD-WAN and other VPN technology domains to intercommunicate

SD-WAN Gateway

- This approach does not require SD-WAN Edges to be placed at each VPN site to achieve interconnectivity

- However, SD-WAN service capabilities such as application-based traffic forwarding over multiple WANs or QoS and Security policy management will not be available at the MPLS VPN sites because they do not have SD-WAN Edges which perform these functions

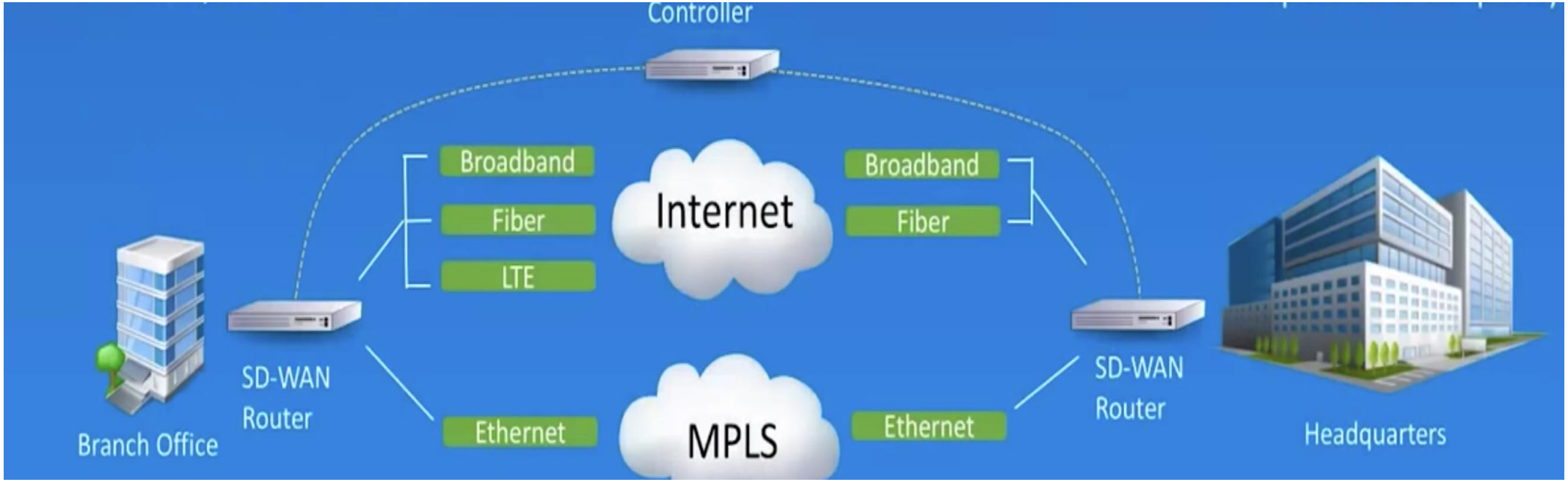
SD-WAN Web Portal

- Subscriber Web Portal is added to the enterprise's existing managed services portal
- It works in conjunction with the service orchestrator to monitor the SD-WAN as a service
- The MSP or CSP typically integrates the Subscriber Web Portal for the SD-WAN managed service into their existing customer portal used for other managed services

SD-WAN Key Characteristics

1. The ability to support multiple connection types, such as MPLS, Last Mile Fiber Optical Network or through high speed cellular networks
   e.g. 4G LTE and 5G wireless technologies
2. The ability to do dynamic application aware path selection, for load sharing and resiliency purposes
3. A simple interface that is easy to configure and manage
4. The ability to support VPNs, and third party services such as WAN optimization controllers, firewalls and web gateways

SD-WAN – Different Transport Mechanisms

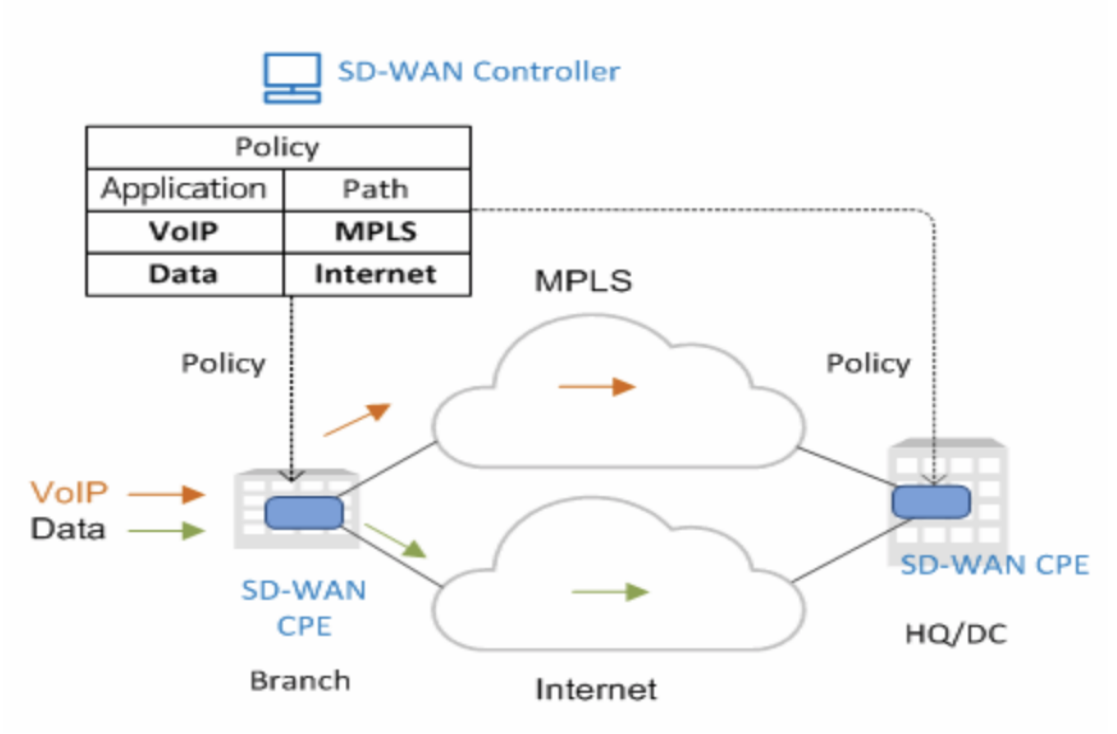


SD-WAN – Dynamic Path Selection

- This feature ensures traffic uses the best path depending on the business need, such as mission-critical and delay-sensitive applications

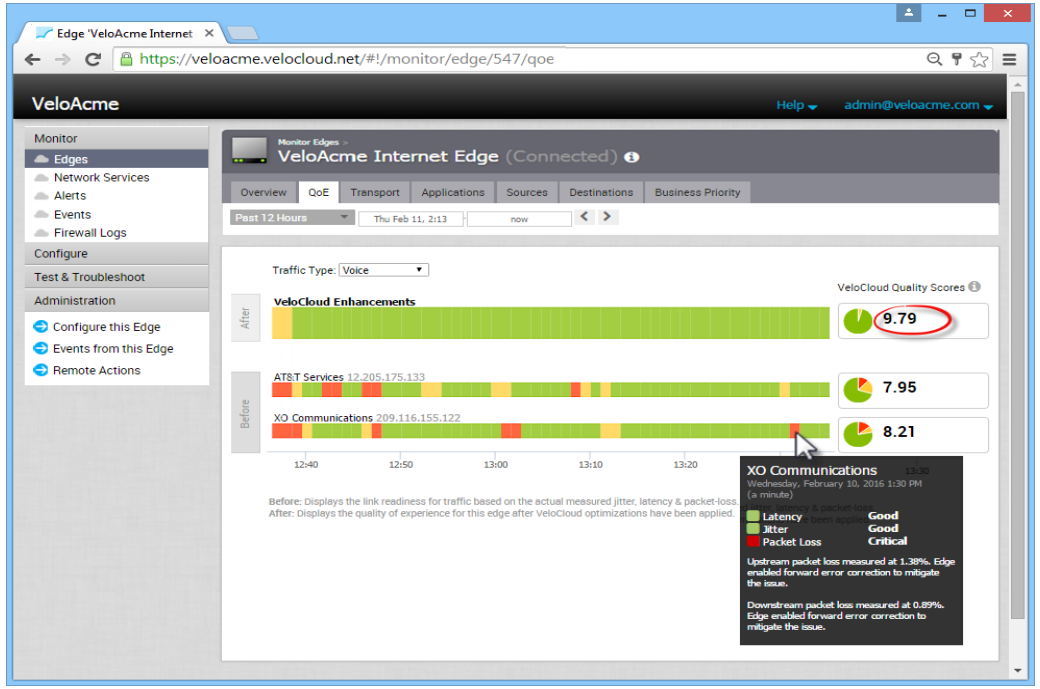
SD-WAN – Dynamic Path Selection

- SD-WAN solution requires a path selection/control solution that allows for each application to dynamically switch their paths in real-time, in response to network conditions, rather than sticking to one particular underlay!



SD-WAN – Simpler Management Compare to Legacy WAN

- GUI provides simpler management, reduced troubleshooting time, mass deployment and update , centralized monitoring and so on.



Source : VeloCloud SD-WAN

SD-WAN – Wan Optimization – Security and Other Services

- WAN optimization accelerates application traffic by overcoming latency and reducing the amount of data traversing the WAN by applying techniques like, deduplication, compression and caching to dramatically increase the amount of available bandwidth

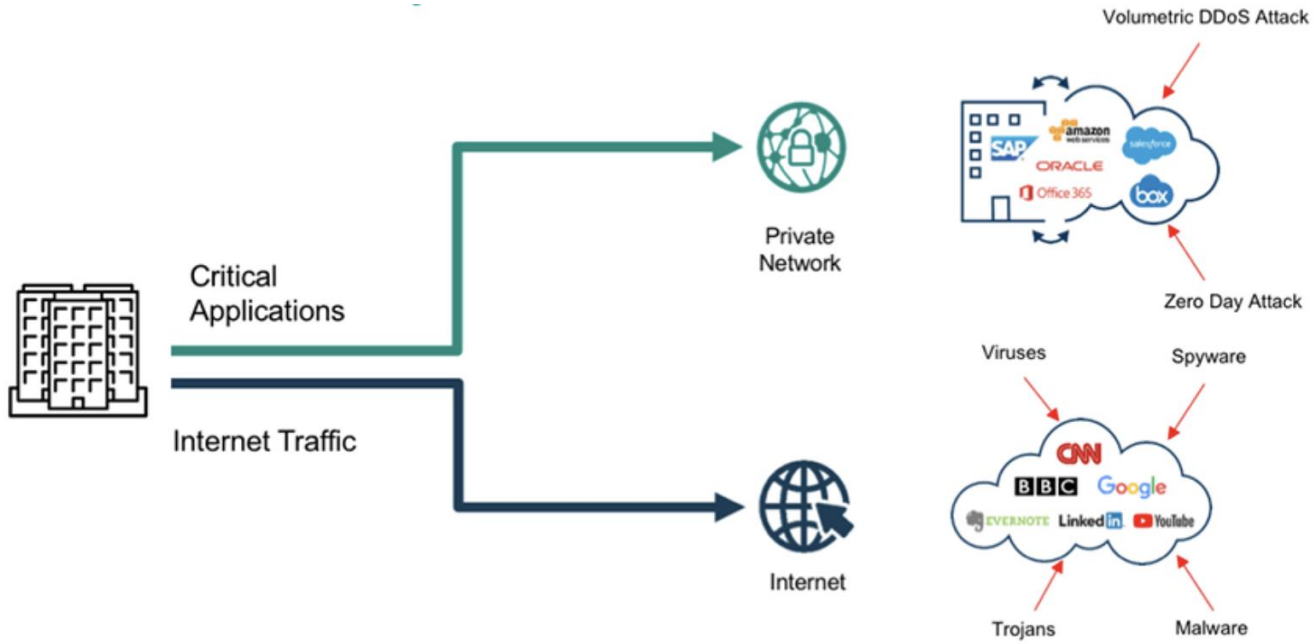
SD-WAN – Wan Optimization – Security and Other Services

- Most SD-WAN implementations offer a way to encrypt your branch-to-branch corporate traffic using IPSEC which protects the data in transit

- Because most SD-WAN vendors offer IPsec, it’s common to think that SD-WANs are inherently secure

SD-WAN – Wan Optimization – Security and Other Services

- It's true that IPsec handles protecting the data as it traverses the network
- But it has no impact on DDOS protection, man-in-the-middle and malware for direct branch-to-cloud traffic
- Centralized security control should be re thought when it comes to SD-WAN security



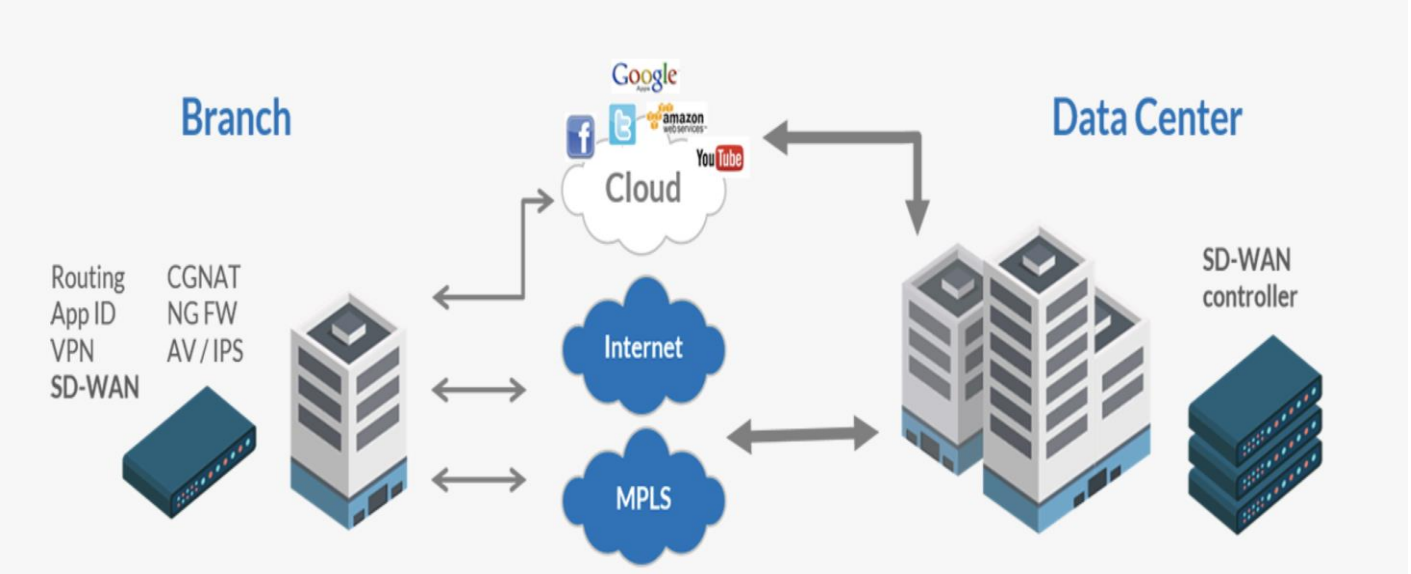
Source : ARYAKA

SD-WAN – Wan Optimization – Security and Other Services

- For example, you still need stateful firewall capabilities between the public Internet and your WAN edge device to grant or deny access

SD-WAN – Wan Optimization – Security and Other Services

- Most NGFWs also comes with a variety of UTM functions, including intrusion detection and prevention (IDS/IPS), quarantining or otherwise deflecting detected malware, and web filtering, which knows about risky Internet sites and prevents your users from visiting them



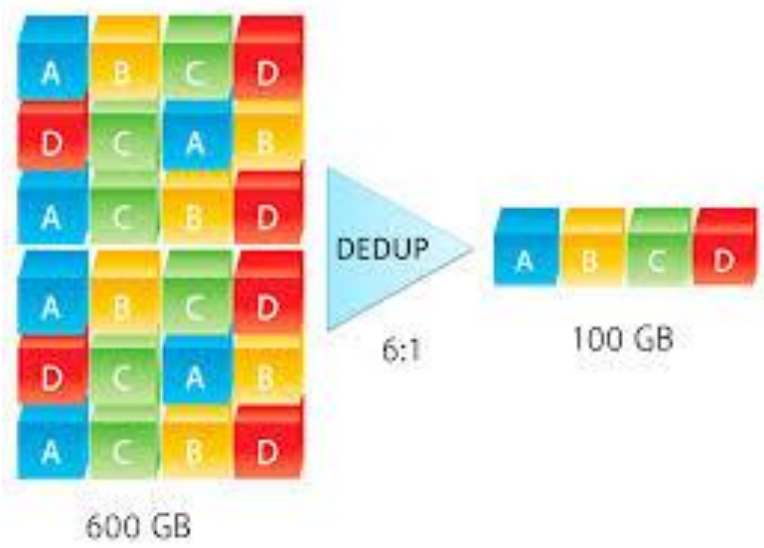
Source : VERSA

SD-WAN – Wan Optimization – Security and Other Services

- Since every branch constitutes a WAN edge with exposure to the Internet, you may need all these capabilities at each branch site!

SD-WAN – Wan Optimization – Deduplication and Compression

- Deduplication analyzes blocks of data, looking for repetition
- It replaces multiple copies of data with references to a single, compressed copy, thereby reducing the amount of capacity needed



SD-WAN – Wan Optimization – Deduplication and Compression

- Data Deduplication (dedupe) provides storage savings by eliminating redundant blocks of data
- Storage capacity reduction is accomplished only when there is redundancy in the data set
- This means the data set must be comprised of multiple identical files or files that contain a portion of data that is identical to the content found in other files

SD-WAN – Wan Optimization – Deduplication and Compression

- Data compression reduces the number of bits required to represent the information
- Compressing large files into smaller bits allows users to store more data and also it makes data transmission much quicker and easier
- Compressed data must be decompressed so that the original data can be extracted and the amount a document is compressed is measured by something called the compression ratio

SD-WAN – Wan Optimization – Deduplication and Compression

- Unlike deduplication, compression is not concerned with whether a second copy of the same block exists, it simply wants to store the most efficient block on the storage

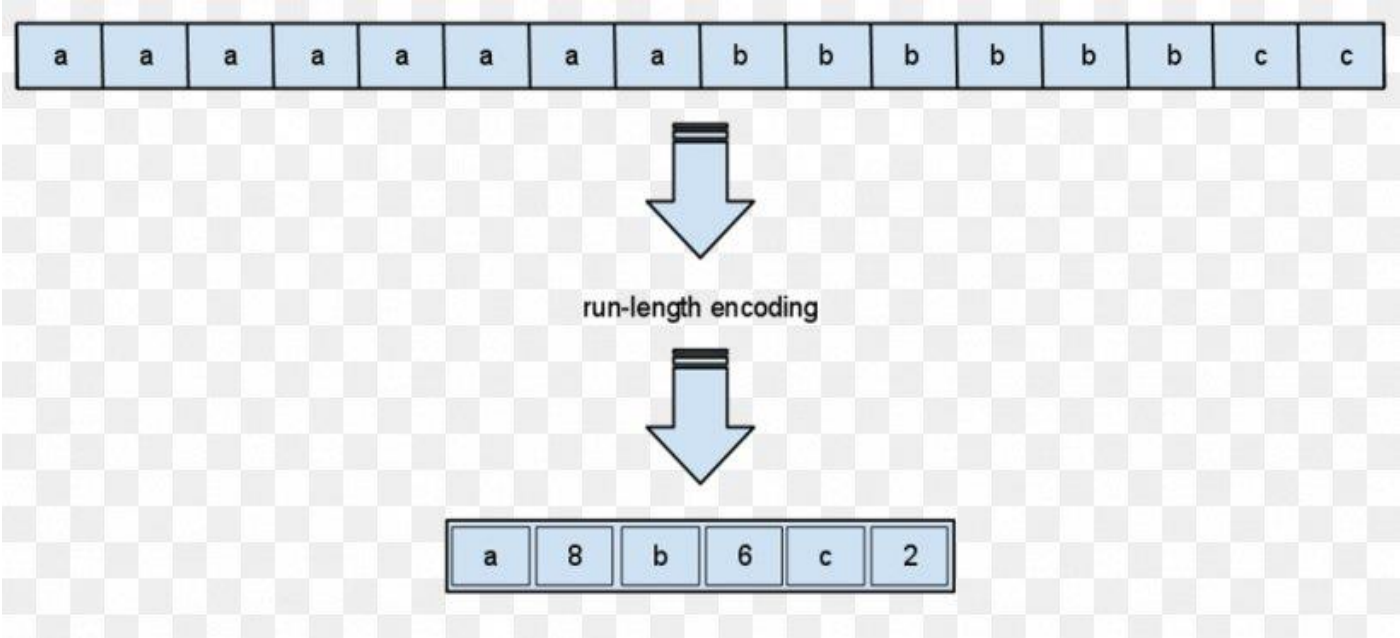
- Examples of common file level compression that we use in our day-to-day lives include MP3 audio and JPG image files

SD-WAN – Wan Optimization – How Data Compression Work?

- Entropy Encoding is one of the techniques for compression
- You might start with a string like: AABCABBCABACBAAACBCCAABAAACBAA
- You might notice that some letters appear more than others - A appears about 2x as much as B and C, and the other letters don't appear at all
- Using that information, you can choose an encoding that represents the characters in the string with less information, e.g., A may be encoded using binary 0, while B and C are assigned 10 and 11respectively. If you were originally using 8 bits per character, that is a big savings

SD-WAN – Wan Optimization – How Data Compression Work?

- Another encoding schema can be Run-length encoding



SD-WAN – Wan Optimization – Security and Other Services

- Packet loss occurs when network congestion or problems in the physical infrastructure cause packets to be lost during transmission

- It’s expressed as a percentage of packets

SD-WAN – Wan Optimization – Security and Other Services

- Packet loss is addressed by some WAN optimization appliances using forward error correction (FEC) that allows receiving stations to automatically regenerate lost packets without requiring transmission

- Let’s have a look at Forward Error Correction

SD-WAN - Forward Error Correction

- For some applications it is necessary to have good error protection
- Sometimes, it will be impossible for the receiver to communicate back with the sender to check for errors in the received packages

SD-WAN - Forward Error Correction

- Some algorithms are made for this kind of situation as for example in a multiple receiver communication

- They use a forward error correction, which is based on the addition of redundant bits over the bit stream of data

SD-WAN - Forward Error Correction

- A simple example of forward error correction is *(3,1) repetition code.* In this example, each bit of data is sent three times and the value or meaning of the message is decided upon majority vote. The most frequently sent bit is assumed to be the value of the message (see table below)

Triplet received	**Interpreted as**
000	0 (error free)
001	0
010	0
100	0
111	1 (error free)
110	1
101	1
011	1

SD-WAN - Good to have capabilities with SD-WAN

- Some of these features might be good to have for some companies and must to have for others depends on the application requirements and the constraints

- Quality of Service , Zero Touch Deployment , Global Coverage , Vendor POC support , Cloud Enablement

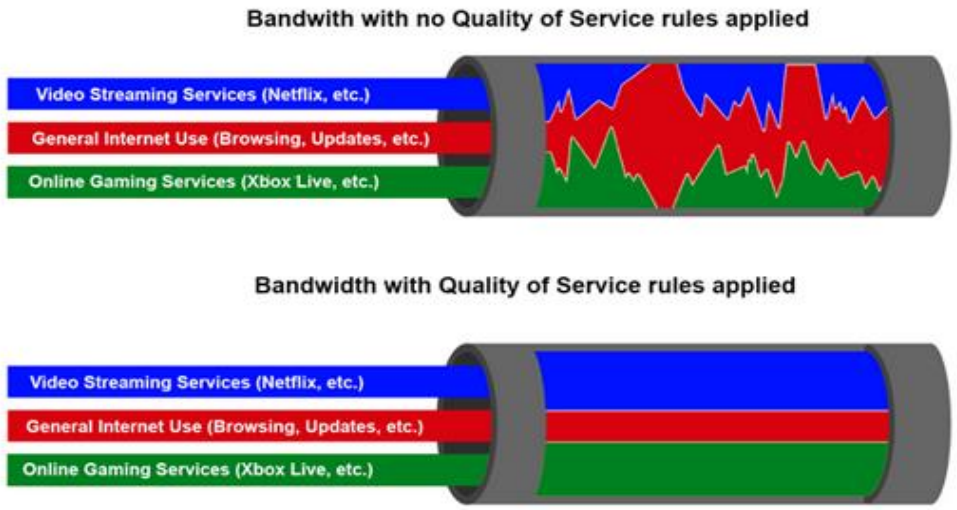
SD-WAN - Quality of Service

- Internet connectivity is one of the cheapest and most widely available bandwidth options

- However, when it comes to building a corporate wide area network (WAN), Internet connectivity is still not seen as a reliable medium for important business data

SD-WAN - Quality of Service

- Quality of service (QoS) refers to the ability of a network to provide higher levels of service using traffic prioritization and control mechanisms



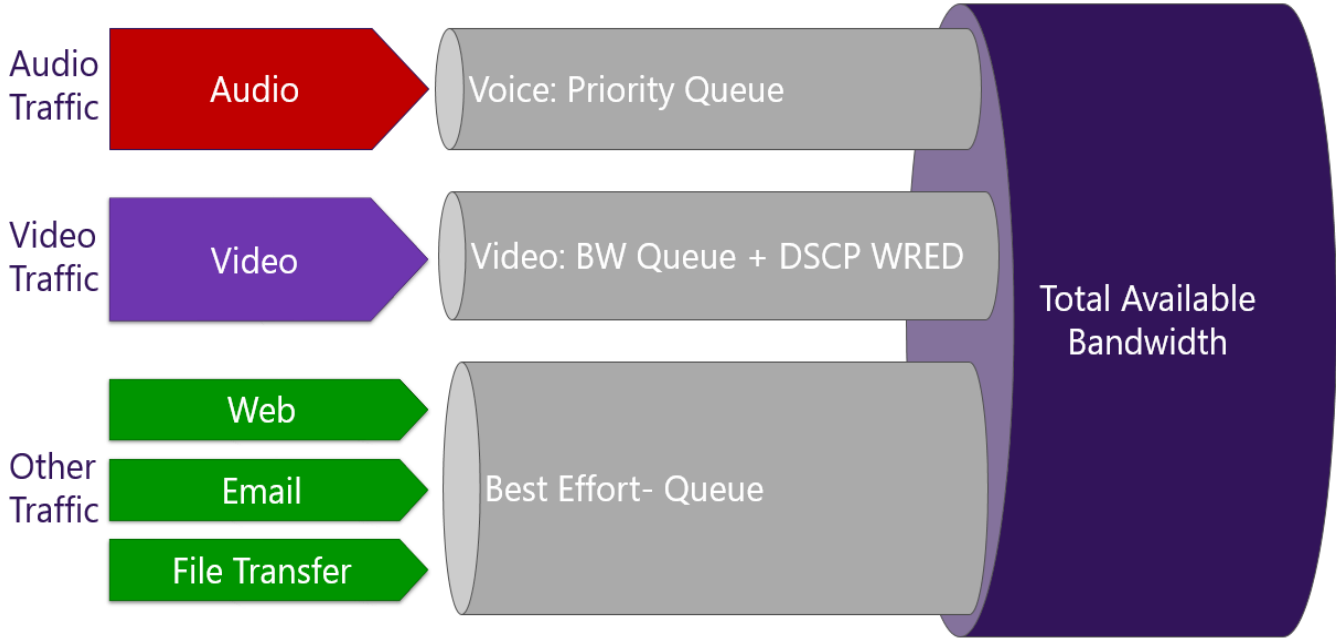
Source : WanDynamics

SD-WAN - Quality of Service

- Some SD-WAN vendors market their Forward Error Correction (FEC) and Dynamic Path Selection/Control features as QOS but they are not QOS mechanisms
- Although these features improve the network performance, they shouldn’t marketed as QOS features!
- Some SD-WAN vendors support Traffic Shaping , Rate Limiting , Policying as QoS features as well

SD-WAN - Quality of Service

- QoS simply prioritization some traffic and punishing others!



SD-WAN - Zero-touch Deployment/Provisioning

- With this capability, IT teams can bring up services without the need to interact with physical equipment, resulting in fast and efficient deployment of services
- ZTP can be found in switches, wireless access points, SD-WAN nodes, NFV-platforms , firewalls and many other networking devices
- Not all ZTP implementations are truly ‘Zero Touch’ though, so sometimes you will also come across terms like ‘minimal touch provisioning’ or ‘one touch provisioning’

SD-WAN - Zero-touch Deployment/Provisioning

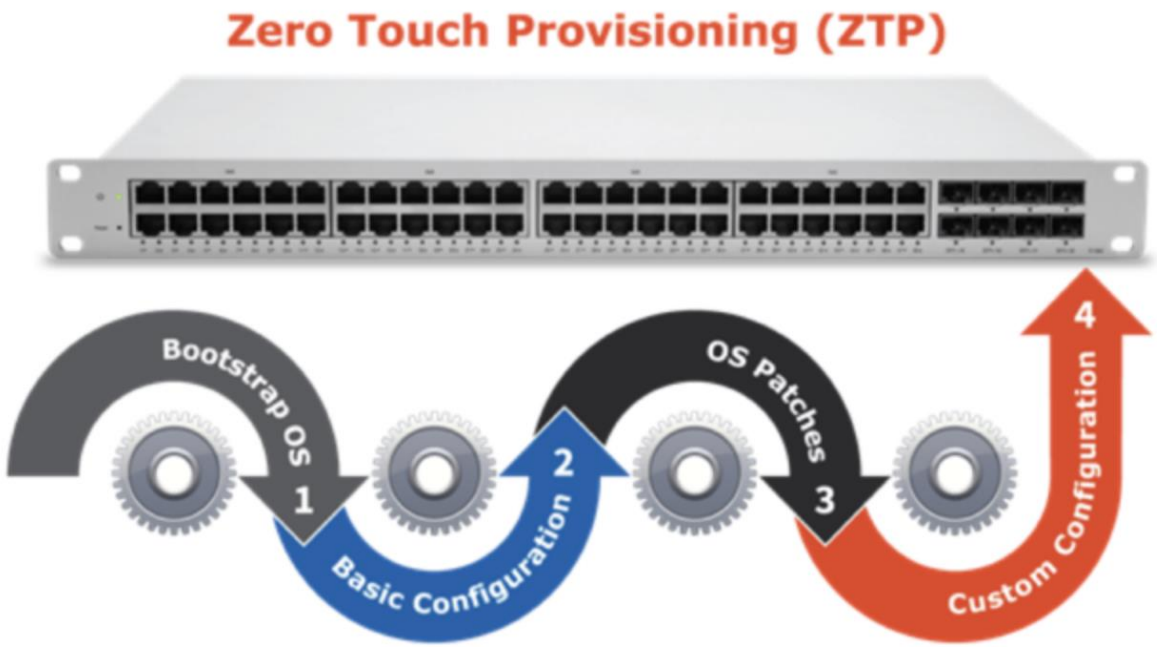
Below steps usually included in a ZTP process

1. Making the device reachable
2. Firmware upgrade
3. Base configuration (DNS, NTP, RADIUS...)
4. Specific configuration (VLANs, interfaces, routing protocols...)

SD-WAN How ZTP Works?

- As the name Zero Touch Provisioning would suggest, the goal is to install a networking appliance somewhere without someone needing to configure it locally

- A new or replacement device can be sent to a site, physically installed and powered up by a locally present employee without IT skills



SD-WAN How ZTP Works?

- Most ZTP implementations are DHCP driven
- Some vendors use ZTP-USB
- Some vendors gives ability to execute scripts (Python or Bash)
- Some vendors retrieves a configuration file via HTTP or TFTP

- Internet connectivity is required in the first place

SD-WAN How ZTP Works?

SD-WAN How ZTP Works?

- More and more vendors are offering a cloud service to support the configuration and ZTP process
- Cisco Meraki, Riverbed, Citrix and Juniper Networks are among those
- All that is required is registering the serial numbers of the devices purchased and the vendor will ensure the devices are correctly registered and visible under your management portal account
- The device can then be fully configured and managed via the cloud

SD-WAN - Global Coverage

- If your business requires international connectivity, you may need to analyze the provider's point-of-presence (POP) coverage to understand the effect on application performance
- Certain providers and vendors operate a significant global network presence that includes specific POPs for both private and internet traffic
- SD-WAN features are focused on application performance, but latency and jitter challenges can arise when deploying international services

SD-WAN Vendor POC Support

- The proof of concept for SD-WAN is an excellent way to understand and verify the capability of an SD-WAN offering

- Some vendors offer demo hardware for a period of time, often with presales resources to assist with the configuration

SD-WAN Cloud Connection

- Some SD-WAN products have the ability to program “cloud breakout” based on applications, allowing direct access to trusted sites (like SalesForce.com), while tunneling traffic to unknown sites to either cloud-based or centrally-based inspection services

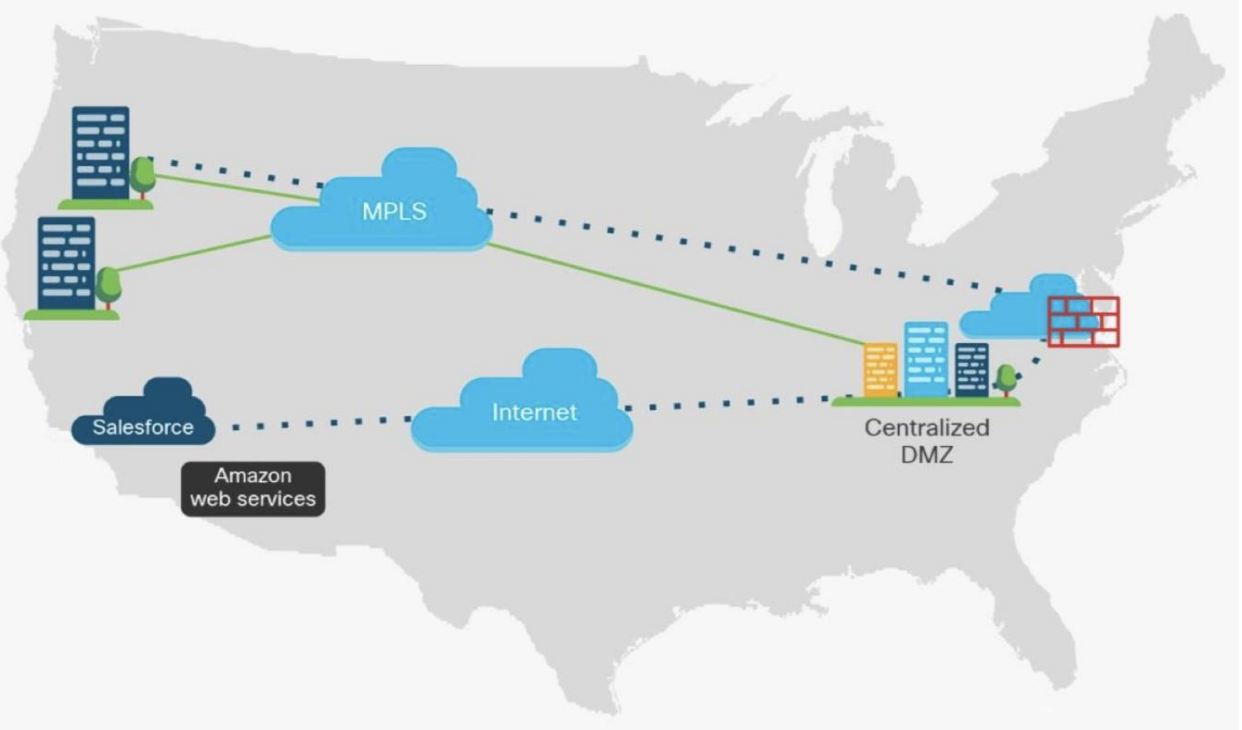
- This ensures improved productivity, minimizes unnecessary inspection of trusted traffic and provides better security than traditional hub-spoke MPLS solutions

SD-WAN Cloud Connection - SAAS

- Enterprises today face major user experience problems for SaaS applications because of networking problems
- The centralized Internet exit architecture can be inefficient and results in poor SaaS performance
- And branch sites are running out of capacity to handle Internet traffic which is a concern because more than 50% of branch traffic is destined to the cloud

SD-WAN Cloud Connection- SAAS

- Common network designs consolidates application and service controls at centralized DMZs and the data centers
- As a result, enterprise traffic destined for the Internet or public clouds must be backhauled through a centralized DMZ facility
- This causes the traffic to trombone or hairpin, creating an inefficient route that increases the distance between the user and the application



Traditional WAN	V.S.	SD-WAN
Months to provision	**Deployment**	**Hours** to provision
On-site technician required to set up new sites		**Zero-touch configuration**
Manual configuration of routers and gateways		**Automatically** deploy traffic rules
Separate management systems for diverse connections and carriers	**Management**	**Unified** user interface to set up / manage diverse connections
High IT expertise required for networks changes and maintenance		**Dynamically self-propagate** configuration changes through network
Limited visibility into network conditions	**Visibility**	**Full visibility** into real-time network conditions
No flexibility for temporary connections	**Flexibility**	**Instantly set up or delete connections** based on real-time demand
Cannot quickly respond to changing bandwidth needs		**Adjust bandwidth quickly** as needed
Not optimized for connectivity to public clouds	**Cloud connectivity**	**Connect in minutes** to multiple clouds and SaaS
Unpredictable public Internet quality can lead to unacceptable latency and packet loss	**Cost**	**Achieve balance between performance and cost** using quality of service (QoS) protocols to differentiate application priorities and
Private / MPLS circuits are expensive and typically require long term contracts		**Pay as you grow**

Cloud Computing, Edge Computing and FOG Computing

Cloud Computing

- Cloud computing is the delivery of computing services including servers, storage, databases, networking, software, analytics, and intelligence over the Internet to offer faster innovation, flexible resources, and economies of scale

- Cloud computing is the delivery of on-demand computing services from applications to storage and processing power typically over the internet and on a pay-as-you-go basis!

Cloud Computing

- Rather than owning their own computing infrastructure or data centers, companies can rent access to anything from applications to storage from a cloud service provider

- One benefit of using cloud computing services is that firms can avoid the upfront cost and complexity of owning and maintaining their own IT infrastructure, and instead simply pay for what they use, when they use it

Cloud Computing – General Characteristics

AGILITY

- The cloud gives you easy access to a broad range of technologies so that you can innovate faster and build nearly anything that you can imagine
- You can quickly spin up resources as you need them–from infrastructure services, such as compute, storage, and databases, to Internet of Things, machine learning, data lakes and analytics, and much more
- You can deploy technology services in a matter of minutes,

Cloud Computing – General Characteristics

ELASTICITY:

- With cloud computing, you don't have to over-provision resources up front to handle peak levels of business activity in the future
- Instead, you provision the amount of resources that you actually need
- You can scale these resources up or down to instantly to grow and shrink capacity as your business needs change

Cloud Computing – General Characteristics

GLOBAL AVAILABILITY:

- With the cloud, you can expand to new geographic regions and deploy globally in minutes
- This is generally true for the Public Cloud Deployments which we will see next

Cloud Computing

- Architecturally Cloud Computing has 3 different models

- Public Cloud, Private Cloud and Hybrid Cloud

Cloud Computing - Public Cloud

- Public clouds are the most common way of deploying cloud computing. The cloud resources (like servers and storage) are owned and operated by a third-party cloud service provider and delivered over the Internet

- The service may be free or a subscription-based offering charged based on the computing resources consumed

Cloud Computing - Public Cloud

- The computing functionality may range from common services such as email, apps and storage to the enterprise-grade OS platform or infrastructure environments used for software development and testing

Cloud Computing – Public Cloud

Characteristics of Public Cloud:

- Lower costs - No need to purchase hardware or software, and you pay only for the service you use
- No maintenance - Service provider provides the maintenance
- Near-unlimited scalability - On-demand resources are available to meet your business needs
- High reliability - A vast network of servers ensures against failure

Cloud Computing – Public Cloud

Concerns with Public Cloud:

- The total cost of ownership (TCO) can rise exponentially for large-scale usage, specifically for midsize to large enterprises
- Low visibility and control into the infrastructure, which may not suffice to meet regulatory compliance.

Cloud Computing – Private Cloud

- Private Cloud refers to the cloud solution dedicated for use by a single organization

- The computing resources are isolated and delivered via a secure private network, and not shared with other customers

Cloud Computing – Private Cloud

- **General Characteristics of Private Cloud:**

- Dedicated and secure environments that cannot be accessed by other organizations

- Flexibility to transform the infrastructure based on ever-changing business and IT needs of the organization

Cloud Computing – Private Cloud

It is mostly suitable for:

1. Highly regulated industries and government agencies
2. Technology companies that require strong control and security over their IT workloads and the underlying infrastructure
3. Large enterprises that require advanced data center technologies to operate efficiently and cost-effectively
4. Organizations that can afford to invest in high performance and availability technologies

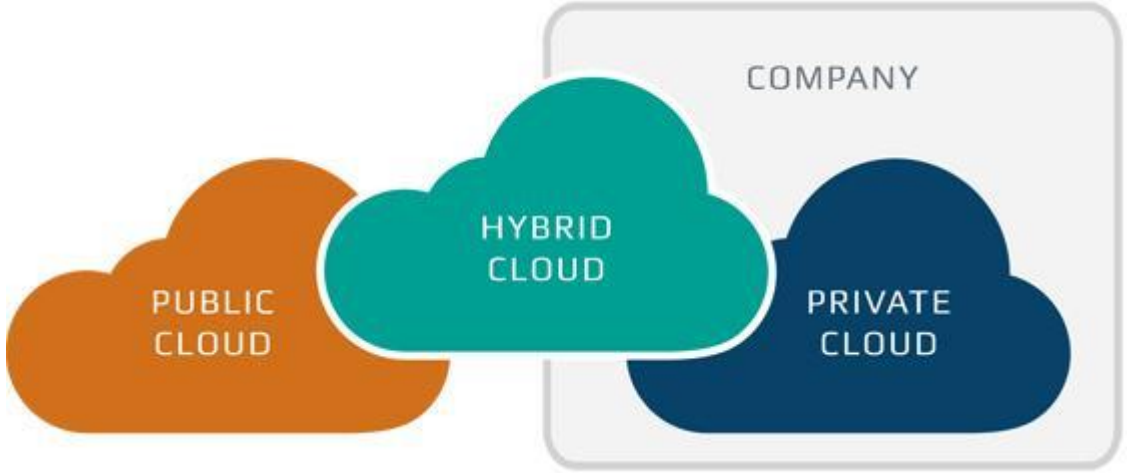
Cloud Computing – Private Cloud

Limitations of Private Cloud:

- Expensive solution with a relatively high total cost of ownership as compared to public cloud alternatives for short-term use cases
- The infrastructure may not offer high scalability to meet unpredictable demands if the cloud data center is limited to on-premise computing resources

Cloud Computing – Hybrid Cloud

- This is a cloud infrastructure environment that is a mix of public and private cloud solutions
- Applications and data workloads can share the resources between public and private cloud deployment based on organizational business and technical requirements



Cloud Computing – Hybrid Cloud

- Often called “the best of both worlds,” hybrid clouds combine on-premises infrastructure, or private clouds, with public clouds so organizations can get the advantages of both

Cloud Computing – Hybrid Cloud

Characteristics of Hybrid Cloud:

- It provides flexibility so you can take advantage of additional resources in the public cloud when you need them
- It is cost effective with the ability to scale to the public cloud, you pay for extra computing power only when needed
- Organization can maintain a private infrastructure for sensitive assets

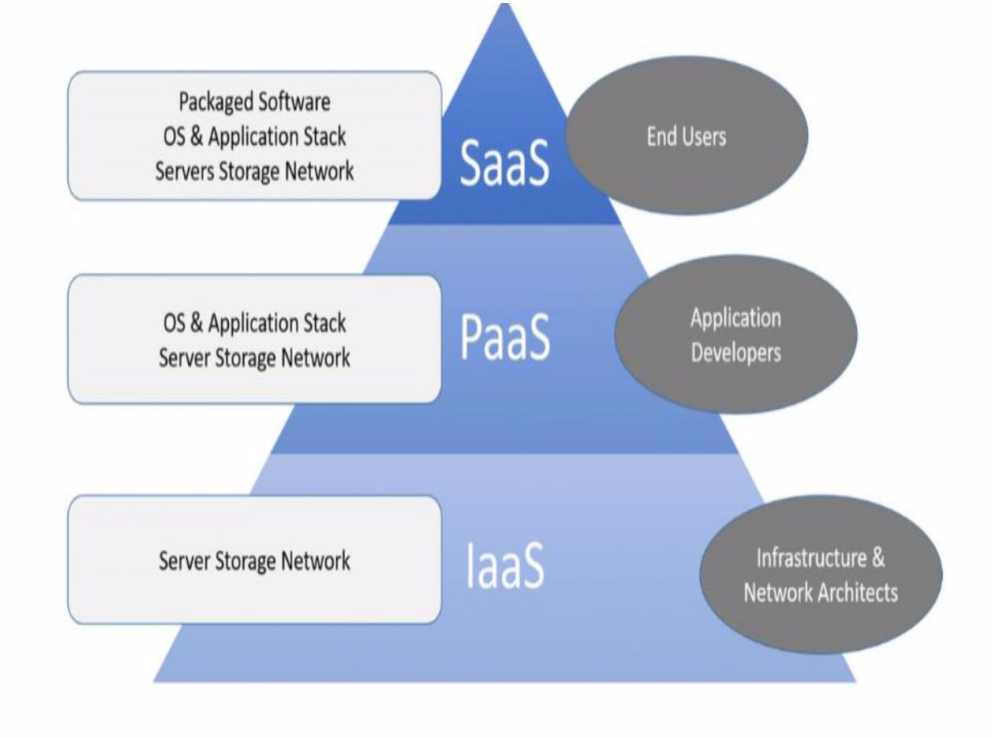
Cloud Computing – Private vs. Public vs. Hybrid Cloud

Design Requirement	Private Cloud	Public Cloud	Hybrid Cloud
Scalablability	Least Scalable because it can scale only with the internal hosted resources	Highly Scalable It can scale up or down based on the requirements	Modarate since it offers both Private and Pubkic Cloud
Security	Most Secure	Least Secure	Modarate
Who use it	Single Organization	Shared Resources	Either single or shared resources as it offers both
Cost	Short term most costly long term depends	Short term least costly long term deper	Short term cost effective than private cloud
Flexibility and Control	Greater Control and Flexibility	Least Control and Flexibility	Moderate as it offers both Private and Public Cloud options

Cloud Computing – Different Cloud Services

- The three main types of cloud computing include Infrastructure as a Service, Platform as a Service, and Software as a Service

- Each type of cloud computing provides different levels of control, flexibility, and management so that you can select the right set of services for your needs



Cloud Computing – Different Cloud Services - IaaS

Infrastructure as a Service (IaaS):

- This service provides the infrastructure, such as Servers, Operating Systems, Virtual Machines, Networks, and Storage etc. on rent basis

- Example: Amazon Web Service, Microsoft Azure

Cloud Computing – Different Cloud Services - PaaS

Platform as a Service (PaaS):

- This service is used in developing, testing and maintaining software's. PaaS is same as IaaS but also provides additional tools such as DBMS, BI services etc.

- Example: Oracle Cloud Platform (OCP), Red Hat OpenShift, Google App Engine

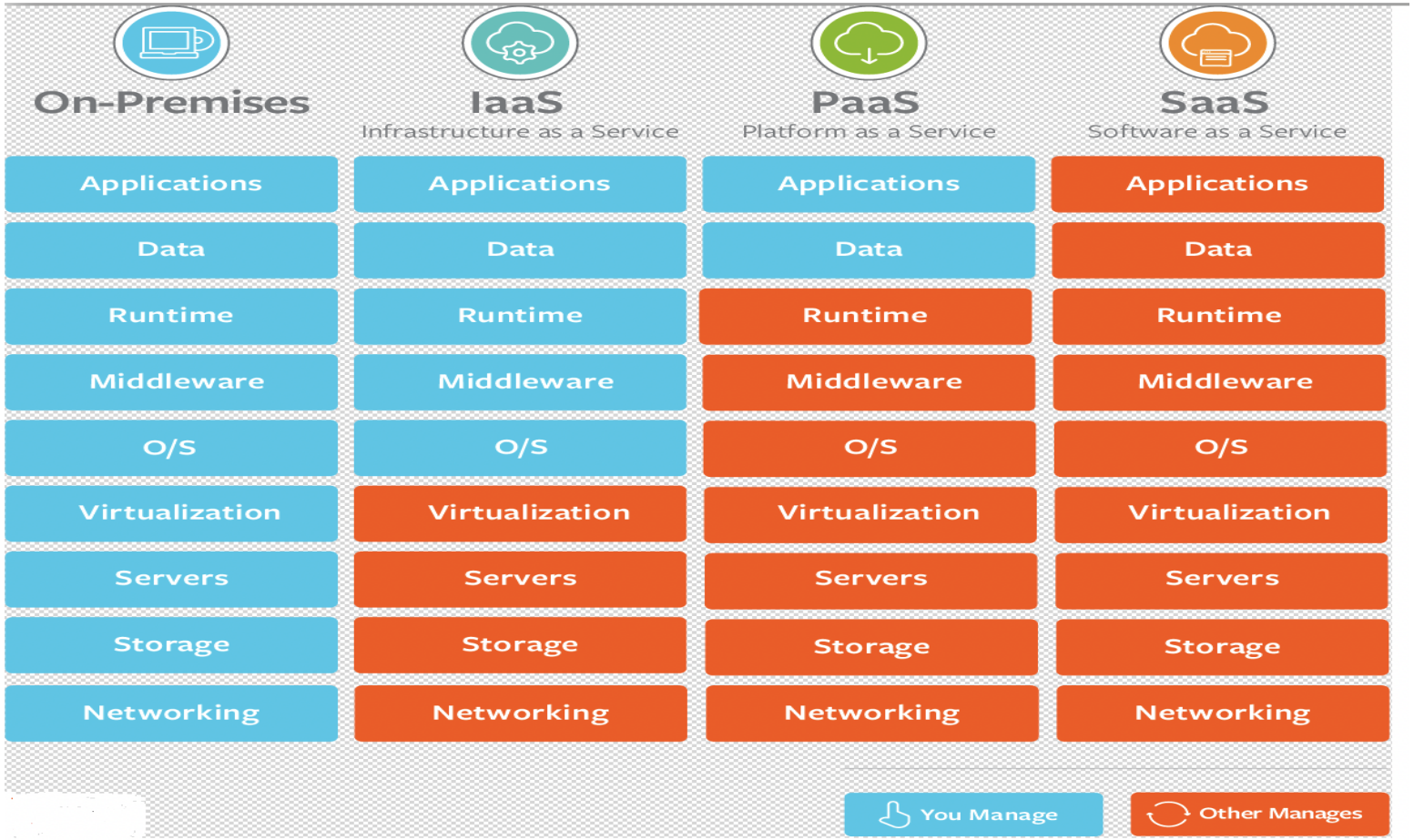
Cloud Computing – Different Cloud Services - SaaS

Software as a Service (SaaS):

- This service makes the users connect to the applications through the Internet on a subscription basis

- Example: Google Applications, Salesforce

Cloud Computing – IaaS vs. PaaS vs. SaaS



Artificial Intelligence (AI) as a Service (AIaaS)

- Because hardware, software and staffing costs for AI can be expensive, many vendors are including AI components in their standard offerings or providing access to artificial intelligence as a service (AIaaS) platforms

- AIaaS allows individuals and companies to experiment with AI for various business purposes and sample multiple platforms before making a commitment

Artificial Intelligence (AI) as a Service (AIaaS)

- **Popular AI cloud offerings include the following:**

- Amazon AI
- IBM Watson Assistant
- Microsoft Cognitive Services
- Google AI

Edge Computing

- Edge computing is a networking philosophy focused on bringing computing as close to the source of data as possible, in order to reduce latency and bandwidth usage

- In a simpler term, edge computing means running fewer processes in the cloud and moving those processes to local places, such as on a user's computer, an IoT device, or an edge server

Edge Computing

- Bringing computation to the network's edge minimizes the amount of long-distance communication that has to happen between a client and server

- It is important to understand that the edge of the network is geographically close to the device, unlike origin servers and cloud servers, which can be very far from the devices they communicate with

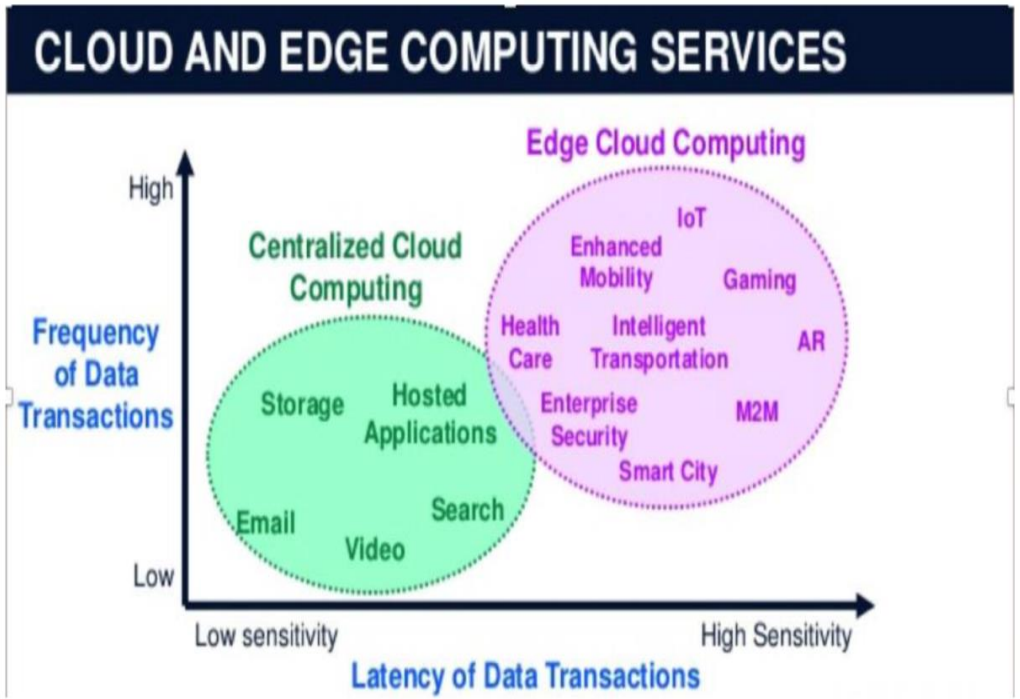
Edge Computing

- Cloud computing offers significant amount of resources (e.g., processing, memory and storage resources) for the computation requirement of mobile applications

- However, gathering all the computation resources in a distant cloud environment started to cause issues for applications that are latency sensitive and bandwidth hungry

Edge Computing

- Akamai, CloudFront, CloudFlare and many other Edge Computing Providers provide edge services like WAF, Edge Applications, Serverless Computing, DDos Protection, Edge Firewall etc.



Fog Computing

- Both the Fog and Edge Computing are totally concerned and looks into the computing capabilities to be done locally rather than pushing it to the Cloud

- Overall reason of having Fog computing is to reduce delay and bandwidth requirement from the network

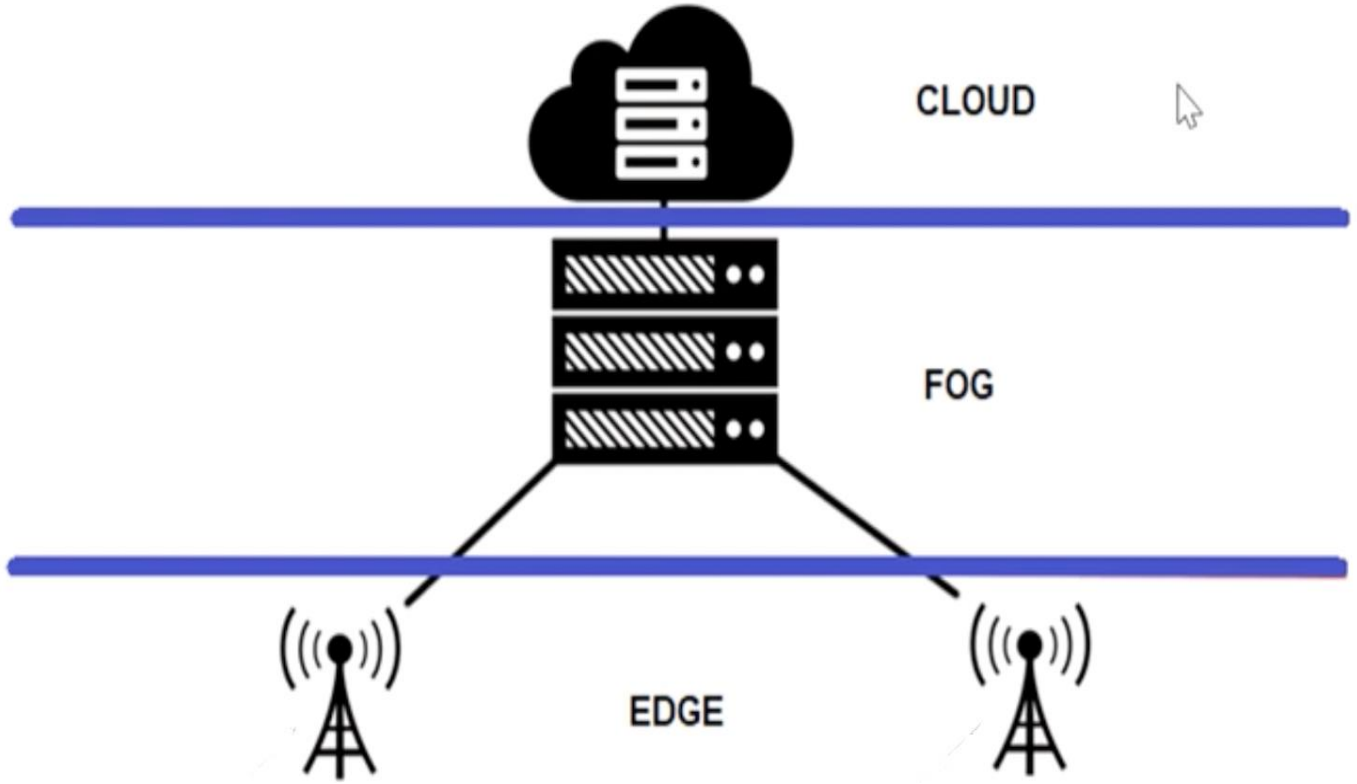
Fog Computing

- Most of the Fog Computing use cases came from the IOT deployments
- Industrial Automation, Intelligent Transportation, Smart Grid etc.
- Edge Computing is heavily discussed with 5G

Fog Computing vs. Edge Computing

- For the real-time applications , having computing resources closer to the source provides faster processing

- Main difference between Fog Computing and Edge Computing is at where the data processing takes place

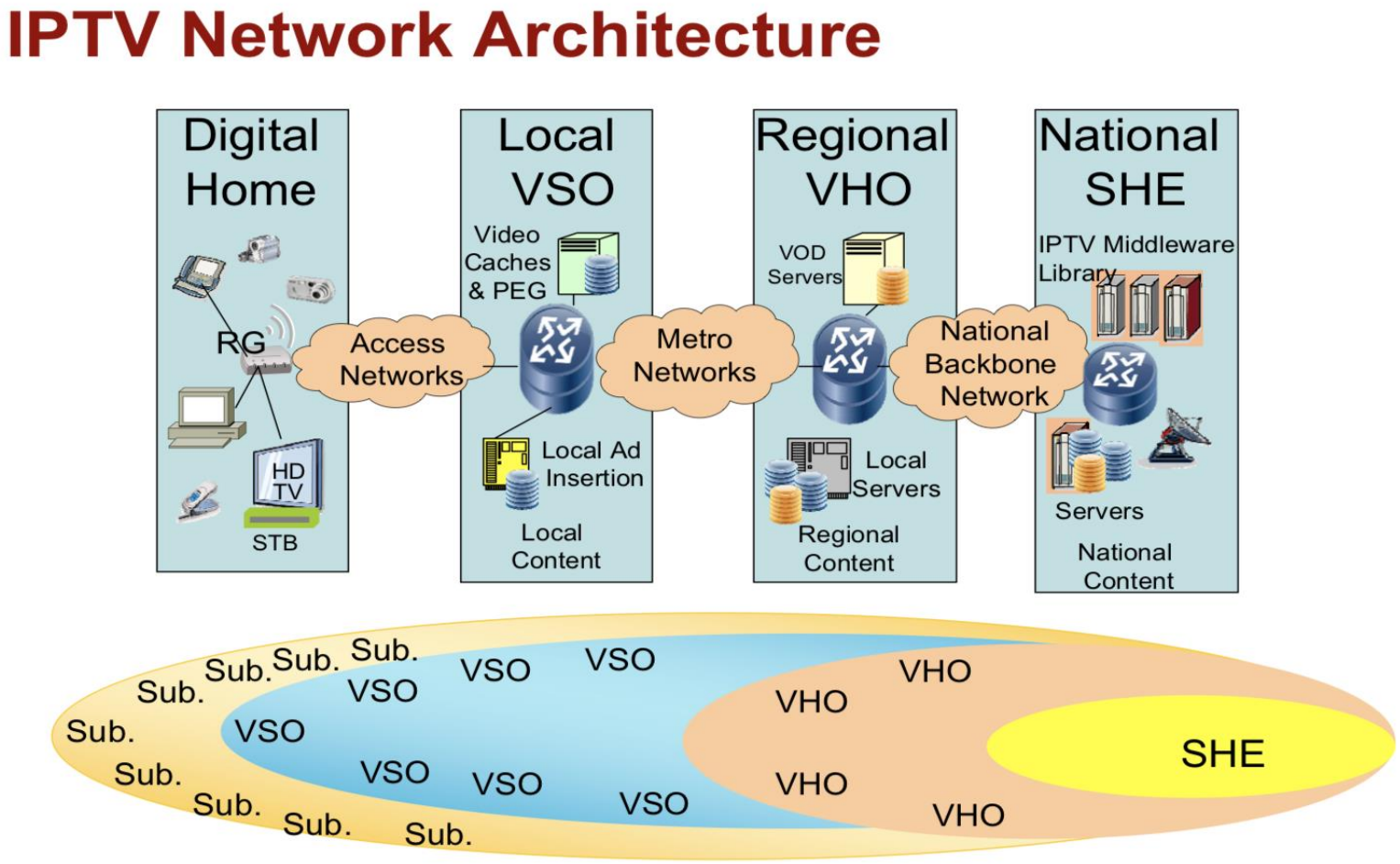


What is IPTV?

- Delivering digital TV services over IP networks by using copper, fiber, wireless and HFC infrastructure

- SP IPTV service is delivered over SP's private network
  - This includes FTTx , xDSL , Wireless , HFC and so on

IPTV Network Architecture



Content Delivery Networks - CDN

CDN – Content Delivery Networks

- Content Delivery Network companies replicate content caches close to large user population
- They don’t provide Internet access or Transit Service to customers or ISPs, but distribute the content of the Content Providers
- Let’s first understand some fundamental businesses such as Content Provider and OTT (Over the Top) Providers

CDN – What does Content Provider do?

- Content Providers are defined as companies that provide actual content to consumers

- There are two types of Internet sources: Eyeballs and Content
  - **Eyeballs**: refers to actual users
  - **Content**: refers to data which the users are interested in

- These two terms are used in the networking communities in the standard bodies (IETF, IEEE etc.) and at the events such as NOG, RIPE and IETF meetings

CDN – What does Content Provider do?

- Search companies (Bing, Google, Yandex, Baidu), TV stations (ABC News, BBC, CNN), video providers (YouTube, Netflix), online libraries and E-Commerce websites all are Content Providers. Content Providers are commonly referred as OTT (Over the Top) Providers

CDN – What does Content Provider do?

- Content Providers have a direct relationship with billions of customers
- Customers pay for ISPs and Content Provider’s services
- Content Providers are not affected by the regulations.
- This is a big debate between Service Providers and Content Providers. All of these regulations lead the Content Providers to become the largest companies in the world

CDN – What does Content Provider do?

6 out of top 10 largest companies in the world are Content Providers

Public / Private Internet Companies, Ranked by Market Valuation (5/29/18)

Rank 2018	Company	Region	Market Value ($B) 5/29/13	5/29/18
1)	Apple	USA	$418	$924
2)	Amazon	USA	121	783
3)	Microsoft	USA	291	753
4)	Google / Alphabet	USA	288	739
5)	Facebook	USA	56	538
6)	Alibaba	China	--	509
7)	Tencent	China	71	483
8)	Netflix	USA	13	152
9)	Ant Financial	China	--	150
10)	eBay + PayPal*	USA	71	133

CDN – What does Content Provider do?

- Google, Netflix, Facebook, Microsoft and almost all other big Content Providers have their own CDN networks and deploy their cache engines widely inside ISP's and/or IXP networks to be closer to their customers
- Large Content Providers have global networks
- If Google were an ISP, it would be the second largest carrier in the planet

CDN – What is OTT?

- Over the Top is a term used to refer to Content Providers

- So, when you hear Over the Top Providers, they are Content Providers. Content can be any application, any service such as Instant messaging services (Skype, WhatsApp), streaming video services (YouTube, Netflix, Amazon Prime), voice over IP and many other voice or video content type

CDN – Advantages of CDN

- Content Distribution Networks reduce latency and increase service resilience (Content is replicated to more than one location)

- More popular contents are cached locally and the least popular ones can be served from the origin

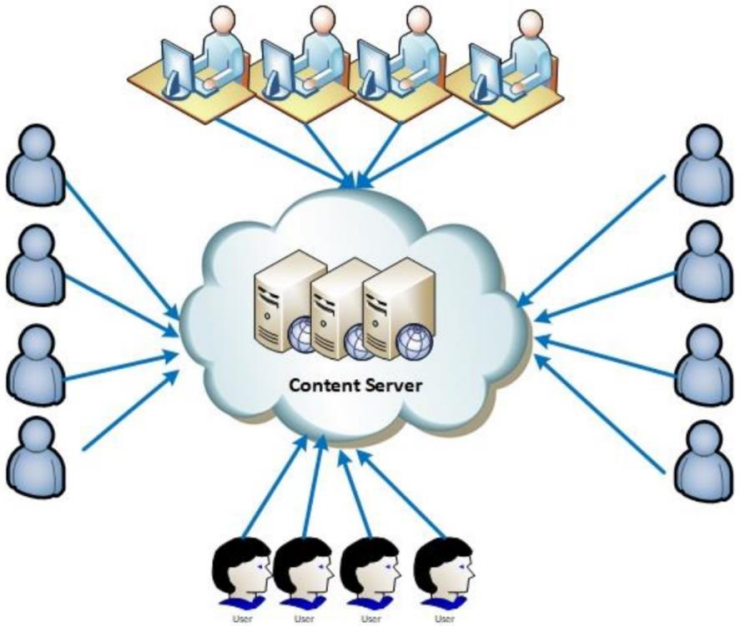
CDN – Advantages of CDN

- Before CDNs, the content were served from the source locations which increased latency, thus reduced throughput

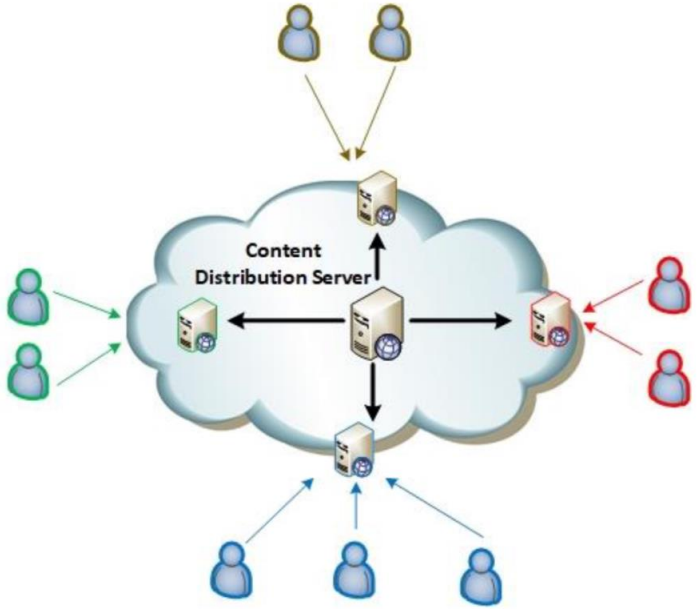
- Content were delivered from the central site. User requests were reaching to the central site where the source was located

CDN – Before CDN – After CDN

BEFORE CDN



AFTER CDN



CDN – Who is doing CDN Business?

- Amazon, Akamai, Limelight, Fastly and Cloudflare are the largest CDN providers which provide services to different Content Providers all over the world

- Also, some major Content Providers such as Google, Facebook, Netflix, etc. prefer to build their own CDN infrastructures and become large CDN providers

CDN – Where CDN Providers Deploy their servers?

- CDN providers have servers all around the world
- These servers are located Inside the Service Providers networks and the Internet Exchange Points
- They have thousands of servers and they serve huge amount of Internet content. CDNs are highly distributed platforms

CDN – How it works?

- To improve user experience and lower transmission costs, large companies set up servers with copies of data in strategic geographic locations around the world

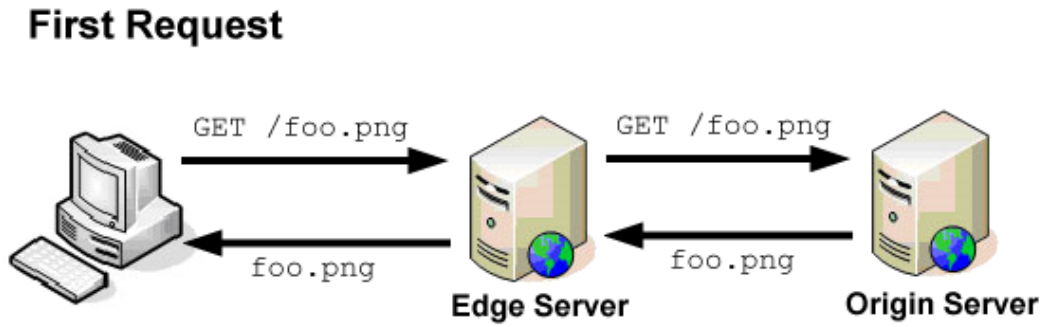
- This is called a CDN, and these servers are called edge servers, as they are closest on the company's network to the end-user

CDN – How it works?

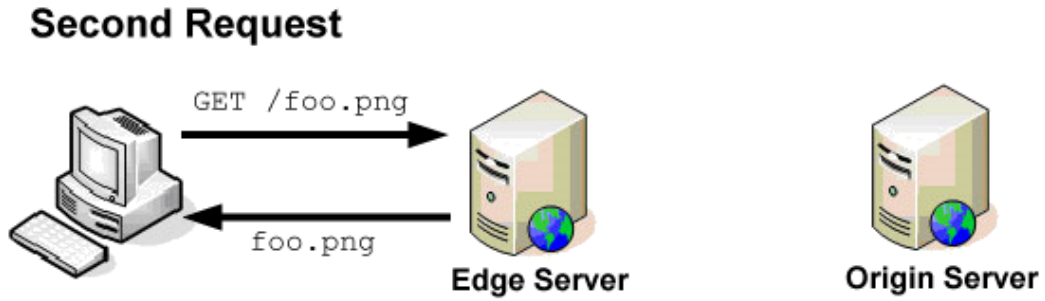
- Edge servers are proxy caches that work in a manner similar to the browser caches
- When a request comes into an edge server, it first checks the cache to see if the content is present
- If the content is in cache and the cache entry hasn't expired, then the content is served directly from the edge server

CDN – How it works?

- If content is not in the cache or the cache entry has expired, then the edge server makes a request to the origin server to retrieve the information



- When the edge server receives the response from the origin server, it stores the content in cache based on the HTTP headers of the response



CDN – Request Routing (Proximity Based Routing)

- Redirecting end users to the optimal edge server, based on the some attributes such as network utilization , end user perceived latency , server load etc. is critical issue and it is called as Request Routing

- Request Routing deals with two issues:
  - Server Selection Mechanism
  - Server Redirection Mechanism

CDN – Request Routing – Server Selection Mechanism

- The mechanism determines the optimal server for an end user

- A server selection algorithm may use a set of metrics, such as network utilization, user perceived latency, network distance, and server load

CDN – Request Routing – Server Selection Mechanism

- Because the nearest server is commonly considered to best serve end users, end user location is typically used as the primary server selection mechanism in request routing

- In practice, most CDN servers simply obtain the end user location from the source IP address of the incoming CDN request

CDN – Request Routing – Server Redirection Mechanisms

- The mechanism informs the end user about the optimal surrogate server selected by the server selection mechanism
- Among all server redirecting mechanisms, DNS-based server redirecting is the most popular
- HTTP Redirection, URL Rewriting and Anycast Methods are the alternative methods to DNS Based approach in Server Redirection

CDN – Request Routing – Server Redirection Mechanisms – HTTP Redirection

- HTTP redirection allows a web server to propagate the server selection result to the end user via HTTP headers

- Hence, the end user can be redirected to the optimal server by following the response generated from the Web server

CDN – Request Routing – Server Redirection Mechanisms – HTTP Redirection

- The weakness of HTTP redirection lies in its reliance on support from the server side to the client side

- Moreover, HTTP redirection is not a lightweight solution because an extra round-trip delay is introduced in every HTTP session, and the processing overheads of HTTP are non-trivial

CDN – Request Routing – Server Redirection Mechanisms – URL Rewriting

- In URL rewriting, the origin server rewrites the generated pages URL links in order to indicate the best server
- Following the rewritten responses, the client can be optimally redirected
- The major cost of URL rewriting is the delay for URL-parsing

CDN – Request Routing – Server Redirection Mechanisms – DNS Based Redirection

- Sites using CDNs configured with DNS Unicast use recursive DNS queries which redirect visitors to their closest node

- This process involves setting up an alternate DNS record for their domain utilizing the DNS CNAME record type

CDN – Request Routing – Server Redirection Mechanisms – DNS Based Redirection

- When a website visitor types in the URL or clicks on a link which references the primary web server, the user's configured DNS provider then redirects them to the CDN service which hosts the site's content

- Once the user reaches the CDN, they are then routed to the closest node which is determined by learning their location based on their IP address

CDN – Request Routing – Server Redirection Mechanisms – DNS Based Redirection

- When the browser makes a DNS request for a domain name that is handled by a CDN, the server handling DNS requests for the domain name looks at the incoming request to determine the best set of servers to handle it

- At it's simplest, the DNS server does a geographic lookup based on the DNS resolver's IP address and then returns an IP address for an edge server that is physically closest to that area

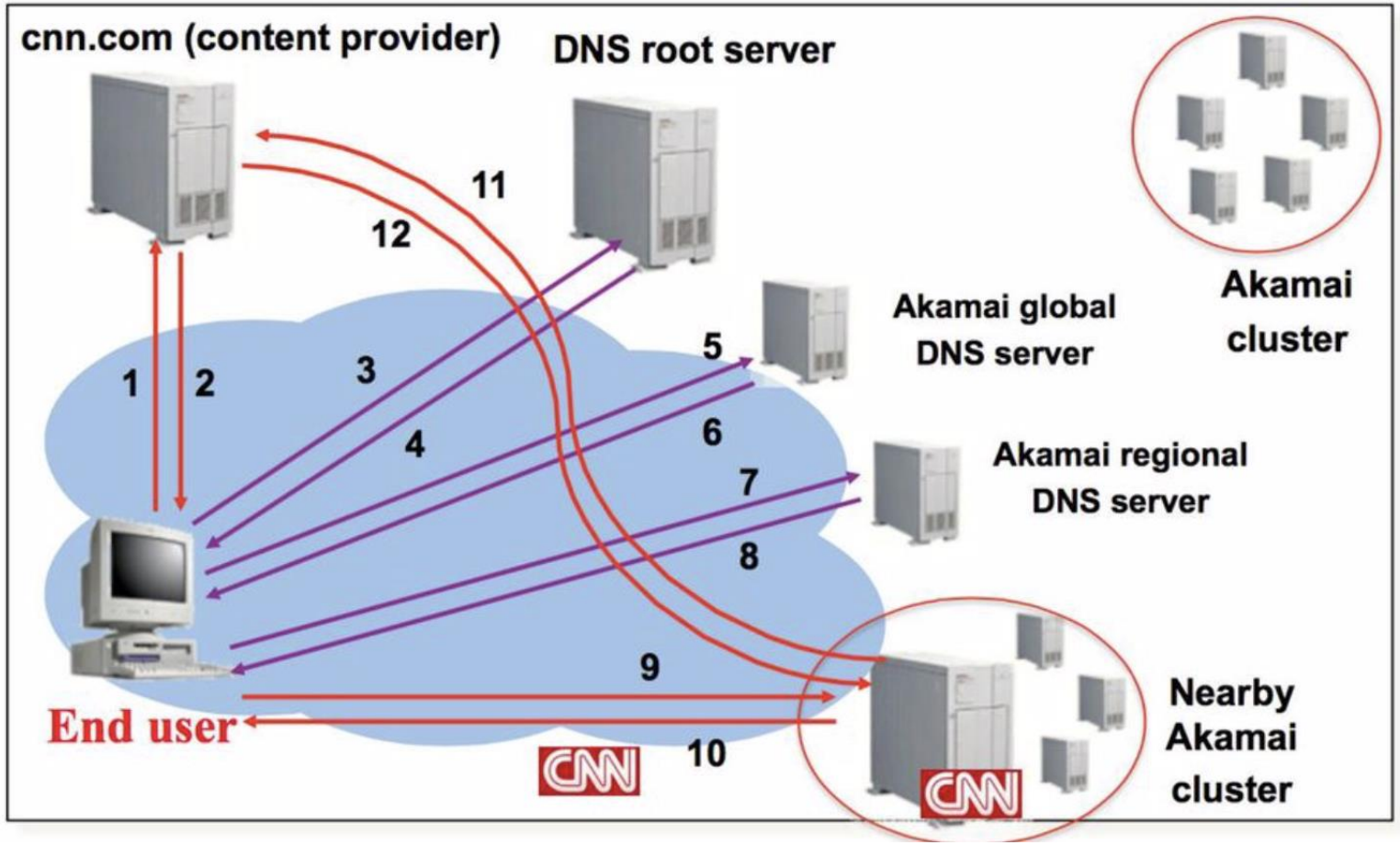
CDN – Request Routing – Server Redirection Mechanisms – DNS Based Redirection

- Companies may optimize their CDNs in other ways as well, for instance, redirecting to a server that is cheaper to run or one that is sitting idle while another is almost at capacity

CDN – DNS Based Redirection Mechanism Problem

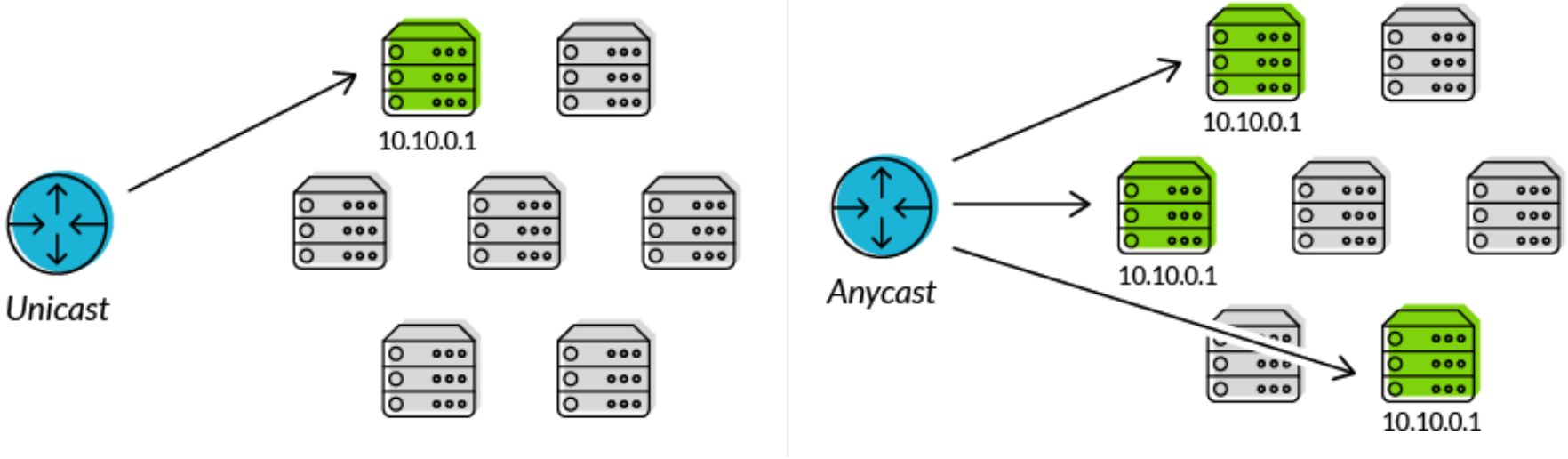
- With this approach, CDN Authoritative DNS Server responds to the IP address of the resolver and not that of the actual client
- For example, a client based in Australia could be configured to use a DNS resolver based in Europe
- In this instance, once the visitor's request has reached the CDN platform, the service will incorrectly assume the originating IP to be in Europe and not Australia sending the client to the incorrect node (EDNS which comes with privacy issue offered as a solution)

CDN – Request Routing – Server Redirection Mechanisms – DNS Based Redirection



CDN – Request Routing – Server Redirection Mechanisms – Anycast

- In this approach, the same IP address is assigned to multiple servers located in a distributed manner



CDN – Request Routing – Server Redirection Mechanisms – Anycast

- When the client sends requests to the IP address, the requests will be routed to the nearest server defined by the routing policy

- With this approach content providers may lose some server selection flexibility Consider a scenario in which Anycast forwards requests to the nearest (yet overloaded) server, by simply respecting a distance-based routing policy

CDN – Request Routing – Server Redirection Mechanisms – Anycast

- CDN service providers who configure their platform with Anycast set a single IP address for all their nodes

- Unlike a DNS Based CDN Redirection, where every node has a unique IP address and recursive DNS routes the client to the closest node, Anycast uses the Border Gateway Protocol (BGP) to route clients using the natural network flow of the Internet

CDN – Request Routing – Server Redirection Mechanisms – Anycast

- BGP is a network level protocol which is used by Internet edge routers to exchange routing and reachability information so that every node on the network, even though it is autonomous, knows the state of their closest network neighbors

- Anycast uses this information to efficiently route traffic based on hop count ensuring the shortest traveling distance between the client and its final destination

CDN – Request Routing – Server Redirection Mechanisms – Anycast

- A CDN configured with Anycast still uses DNS
- The primary difference being that with Anycast only a single IP address is advertised by the CDN provider whereas with DNS based approach each node has a unique IP address
- This CDN routing approach uses the originating client IP instead of the IP of the DNS resolver which ensures the CDN directs the client to the closest possible node

CDN – Request Routing – Server Redirection Mechanisms – Anycast

- Due to its architecture, Anycast offers some advantages over DNS-based request routing
- Most importantly, due to its efficient use of network hops, it allows for faster connectivity
- The complexity in setting up an Anycast solution is also significantly reduced as every node in the network receives a single DNS server configuration

CDN – Request Routing – DNS vs. Anycast

- Anycast approach is considered in a scenarios where less latency , less DNS lookup is required

- DNS Based request routing though, offers more granular server selection criteria, for example server load , POP bandwidth capacity , not only user latency to the server

Wireless Local Area Network Design

WLAN Architecture

There are three different WLAN Architecture:

- Autonomous WLAN architecture
- Centralized WLAN architecture
- Distributed WLAN architecture

WLAN Architecture - Autonomous WLAN Architecture

- For many years, the conventional access point was a standalone WLAN device where control, data and management planes of operation existed and operated on the edge of the network architecture
- These APs are often referred to as fat APs or standalone Aps
- However, the most common industry term for the traditional access point is autonomous AP

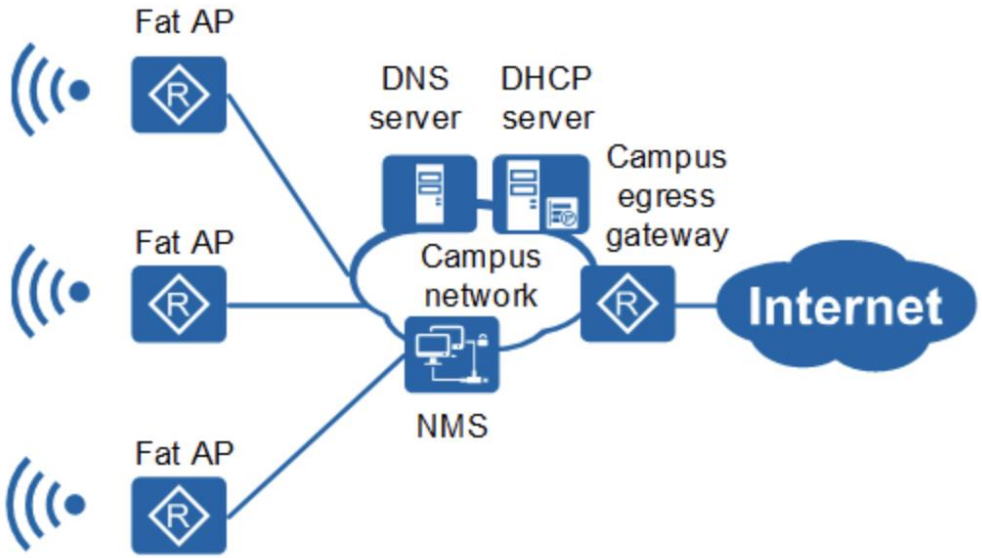
WLAN Architecture - Autonomous WLAN Architecture

- All configuration settings exist in the autonomous access point itself, and therefore, the management plane resides individually in each autonomous AP

- All encryption and decryption mechanisms and MAC layer mechanisms also operate within the autonomous AP

WLAN Architecture - Autonomous WLAN Architecture

- The data plane also resides in each autonomous AP because all user traffic is forwarded locally by each individual access point



WLAN Architecture - Autonomous WLAN Architecture

- Access points operate as layer 2 devices; however, they still need a layer 3 address for connectivity to an IP network
- The BVI is the management interface of an AP
- An autonomous access point typically encompasses both the 802.11 protocol stack and the 802.3 protocol stack

WLAN Architecture - Centralized WLAN Architecture

- This model uses a central WLAN controller that resides in the core of the network

- In the centralized WLAN architecture, autonomous APs have been replaced with controller-based access points, also known as lightweight APs or thin APs

WLAN Architecture - Centralized WLAN Architecture

- Beginning in 2002, many WLAN vendors decided to move to a WLAN controller model where all three logical planes of operation would reside inside the controller

- In a centralized WLAN architecture, the three logical planes exist in a WLAN controller

WLAN Architecture - Centralized WLAN Architecture

- **Management Plane:** Access points are configured and managed from the WLAN controller
- **Control Plane:** Adaptive RF, load balancing, roaming handoffs, and other mechanisms exist in the WLAN controller
- **Data Plane:** The WLAN controller exists as a data distribution point for user traffic

WLAN Architecture - Centralized WLAN Architecture

- Access points tunnel all user traffic to a central controller

- The encryption and decryption capabilities might reside in the centralized WLAN controller or may still be handled by the controller-based APs, depending on the vendor

Centralized WLAN Architecture - WLAN Controller

WLAN Controller has several functions:

- AP Management
- WLAN Management
- User Management
- Device Monitoring
- VLANs
- Security Support
- Captive Portal
- Adaptive RF Spectrum Management
- Layer 3 Roaming Support

Centralized WLAN Architecture - WLAN Controller

There are two types of data-forwarding methods when using WLAN controllers:

- Centralized Data Forwarding
- Distributed Data Forwarding

Centralized WLAN Architecture - WLAN Controller

- **Centralized Data Forwarding:** Where all data is forwarded from the AP to the WLAN controller for processing

- It may be used in many cases, especially when the WLAN controller manages encryption and decryption or applies security and QoS policies

Centralized WLAN Architecture - WLAN Controller

- **Distributed Data Forwarding:** Where the AP performs data forwarding locally may be used in situations where it is advantageous to perform forwarding at the edge and to avoid a central location in the network for all data, which may require significant processor and memory capacity at the controller

Distributed WLAN Architecture

- A recent trend has been to move away from the centralized WLAN controller architecture toward a distributed architecture

- Some WLAN vendors, such as Aerohive Networks, have designed their entire WLAN system around a distributed architecture

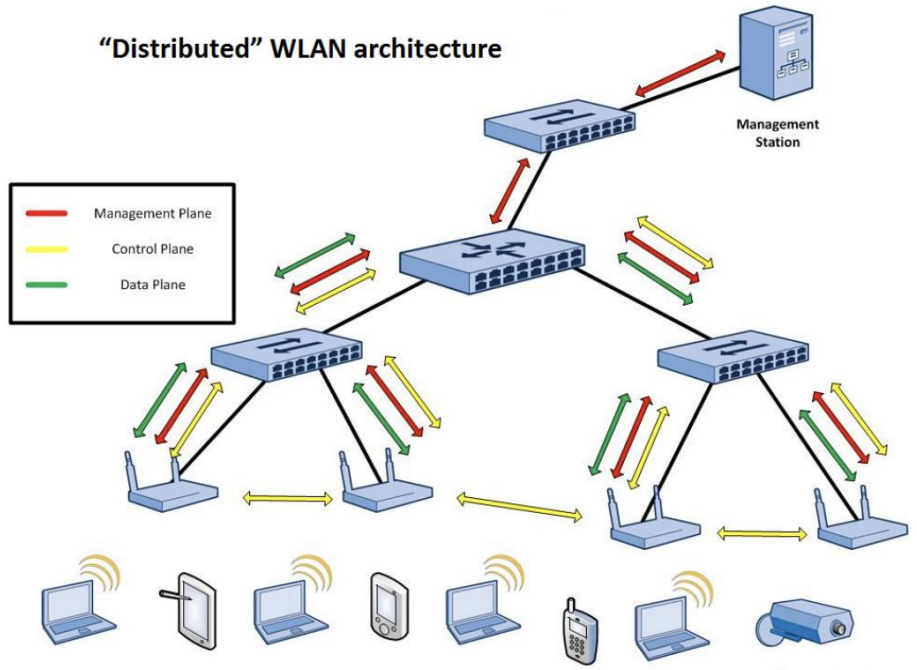
Distributed WLAN Architecture

- Some of the WLAN controller vendors now also offer a distributed WLAN architecture solution, in addition to their controller-based solution

- In these systems, cooperative access points are used, and control plane mechanisms are enabled in the system with inter-AP communication via cooperative protocols

Distributed WLAN Architecture

- A distributed WLAN architecture combines multiple access points with a suite of cooperative protocols, without requiring a WLAN controller



Source: arubanetworks

Distributed WLAN Architecture

- Distributed WLAN architectures are modeled after traditional routing and switching design models, in that the network nodes provide independent distributed intelligence but work together as a system to cooperatively provide control mechanisms

WLAN Design

We will be covering below WLAN Design topics:

- Coverage Design
- Roaming Design
- Channel Design
- Capacity Design

Coverage Design

- When designing a WLAN, probably the first thing that comes to mind will always be the coverage area or zone from which Wi-Fi clients can communicate

- The primary coverage goals for any WLAN are to provide high data rate connectivity for connected clients and to provide for seamless roaming

Coverage Design

- A common mistake is to design a WLAN based solely on an access point's capabilities
- The exact opposite should be considered during the design phase. A proper WLAN coverage design should be based on the perspective of the Wi-Fi clients
- Therefore, a quality received signal for the client is needed to provide high data rate connectivity

Coverage Design – Received Signal

- So what exactly is considered a quality received signal? As shown in the table, depending on the proximity between an AP and a Wi-Fi client, an 802.11 radio might receive an incoming signal anywhere between –30 dBm and the noise floor

Quality	dBm	mW
Very Strong	–30 dBm	1/1,000th of 1 milliwatt
Very Strong	–40 dBm	1/10,000th of 1 milliwatt
Very Strong	–50 dBm	1/100,000th of 1 milliwatt
Very Strong	–60 dBm	1 millionth of 1 milliwatt
Strong	–70 dBm	1 ten-millionth of 1 milliwatt
Fair	–80 dBm	1 hundred-millionth of 1 milliwatt
Weak	–90 dBm	1 billionth of 1 milliwatt
Very Weak	–95 dBm	Noise floor

Coverage Design – Received Signal

- When designing for coverage, the normal recommended best practice is to provide for a –70 dBm or stronger received signal that is well above the noise floor

- In other words, a received signal of –70 dBm or higher is considered to be a quality received signal

Coverage Design – Received Signal

- A received signal of –70 dBm or higher usually guarantees that a client radio will use one of the highest data rates that the client is capable of!

Coverage Design - Signal-to-Noise Ratio (SNR)

- Another reason for planning for –70 dBm coverage is because the received signal of –70 dBm is usually well above the noise floor

- Signal-to-noise ratio (SNR) is an important value because if the background noise is too close to the received signal or the received signal level is too low, data can be corrupted

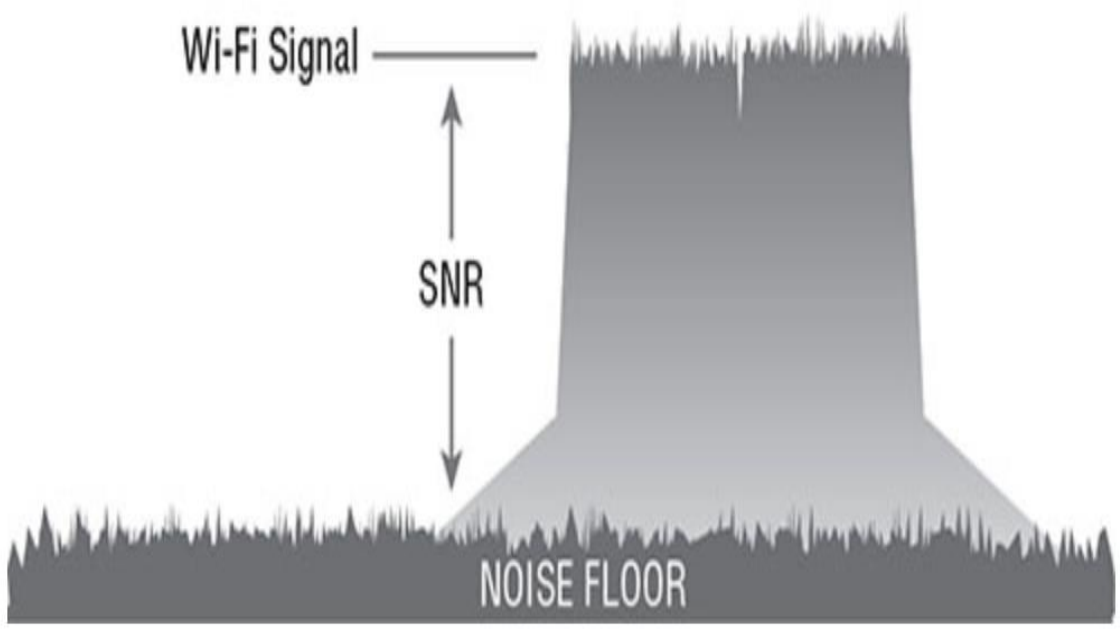
Coverage Design - Signal-to-Noise Ratio (SNR)

- The SNR is not actually a ratio; it is simply the difference in decibels between the received signal and the background noise (noise floor) measured in dBs

- If an 802.11 radio receives a signal of –70 dBm and the noise floor is measured at –95 dBm, the difference between the received signal and the background noise is 25 dB. So, the SNR is 25 dB

Coverage Design - Signal-to-Noise Ratio (SNR)

- In most instances, a received signal of –70 dBm will be 20 dB or higher above the noise floor. In most environments, a –70 dBm signal ensures high rate connectivity, and the 20 dB SNR ensures data integrity



Coverage Design - Signal-to-Noise Ratio (SNR)

- Data transmissions can become corrupted with a very low SNR
- If the amplitude of the noise floor is too close to the amplitude of the received signal, data corruption will occur and result in layer 2 retransmissions
- An SNR of 25 dB or greater is considered good signal quality, and an SNR of 10 dB or lower is considered poor signal quality

Coverage Design - Signal-to-Noise Ratio (SNR)

- An SNR of below 10 dB will likely result in data corruption and retransmission rates as high as 50 percent

- To ensure that frames are not corrupted due to a low SNR, most WLAN vendors recommend a minimum SNR of 20 dB for data WLANs and a minimum SNR of 25 dB for WLANs that require voice-grade communications

Coverage Design - VoWifi

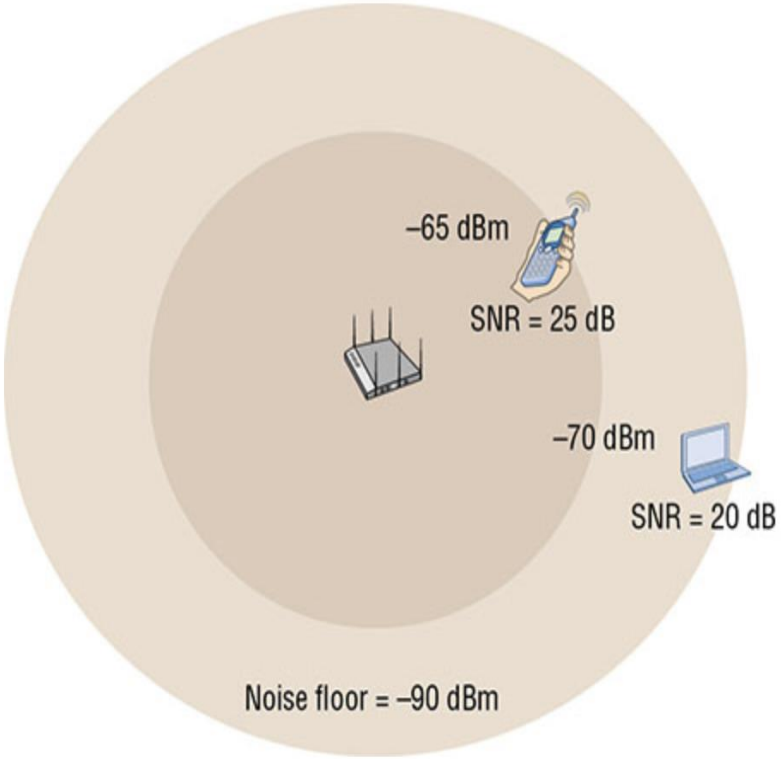
- VoWiFi communications are more susceptible to layer 2 retransmissions than other types of application traffic

- Therefore, when you are designing for voice-grade WLANs, a –65 dBm or stronger signal is recommended so that the received signal is higher above the noise floor

Coverage Design - VoWifi

- As shown in the right figure, even if the noise floor were a very high –90 dBm, the SNR of a –65 dBm received signal for a VoWiFi client would still be 25 dB

- Always check the recommendations of the manufacturer of the VoWiFi client



Coverage Design - VoWifi

- One VoWiFi vendor may state that a –67 dBm signal is sufficient, whereas another vendor may suggest an SNR as high as 28 dB

- When you are designing for voice, SNR is the most important RF metric

Coverage Design – Dynamic Rate Switching

- Will a client device be able to communicate with an AP if the signal drops below –70 dBm?
- The answer is yes, because most client devices can still decode an 802.11 preamble from received signals that are as low as only 4 dB above the noise floor
- As mobile client radios move away from an access point, they will shift down to lower-bandwidth capabilities by using a process known as dynamic rate switching (DRS)

Coverage Design – Dynamic Rate Switching

- Data rate transmissions between the access point and the client stations will shift down or up, depending on the quality of the signal between the two radios

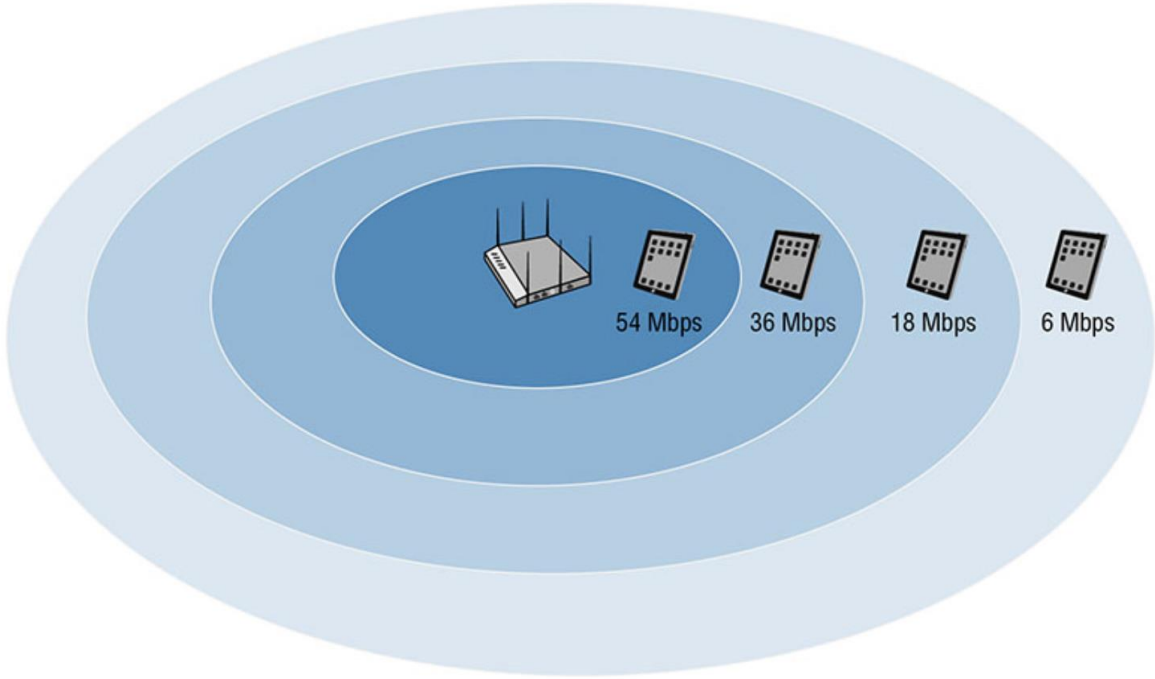
Coverage Design – Dynamic Rate Switching

- There is a direct correlation between signal quality and distance from the AP

- As mobile client stations move farther away from an access point, both the AP and the client will shift down to lower rates that require a less complex modulation and coding scheme (MCS)

Coverage Design – Dynamic Rate Switching

- In the right figure, an 802.11a/g client might connect at 54 Mbps when receiving a –70 dBm signal, but it might shift to transmitting with a lower data rate of 6 Mbps if the signal is much weaker



Coverage Design – Dynamic Rate Switching

- Dynamic rate switching (DRS) is also referred to as dynamic rate shifting, dynamic rate selection, adaptive rate selection, and automatic rate selection

- All these terms refer to a method of speed fallback on a Wi-Fi radio receiver (Rx) as the incoming signal strength and quality from the transmitting Wi-Fi radio decreases

Coverage Design – Dynamic Rate Switching

- The objective of DRS is upshifting and downshifting for rate optimization and improved performance

- From the client's perspective, the lower data rates will provide larger concentric zones of coverage than the higher data rates

Coverage Design – Transmit Power

- A big factor that will affect both WLAN coverage and roaming is the transmit power of the access points
- Although most indoor APs may have full transmit power settings as high as 100 mW, they should rarely be deployed at full power
- This extends the effective range of the access point; however, designing WLANs strictly for range is an outdated concept

Coverage Design – Transmit Power

- WLAN capacity design and reduction of airtime consumption are important design considerations!

- APs at maximum transmit power will result in oversized coverage and not meet your capacity needs

Coverage Design – Transmit Power

- Access points deployed at full transmit power in an indoor environment will also increase the chance of co-channel interference, which can result in unnecessary medium contention overhead

- APs at full power also increase the chance of sticky clients which negatively impact roaming

Coverage Design – Transmit Power

- Typical indoor WLAN deployments are designed with the APs set at about one-fourth to one-third maximum transmit power

- Higher user density environments may require that the AP transmit power be set at the lowest setting of 1 mW

Coverage Design – Transmit Power

- Another consideration is the transmit power of the **CLIENTS**!
- One heavily debated topic is the concept of a balance power link between an AP and a client
- In simpler words, the transmit power settings between an AP and a client are the same
- Very often WLAN clients transmit at higher power levels than indoor access points

Coverage Design – Transmit Power

- The transmit power of many indoor APs may be 10 mW or less due to high-density design needs
- However, most clients, such as smartphones and tablets, may transmit at fixed power of 15 mW or 20 mW
- Because clients often transmit at a higher power than the APs and because clients are mobile, co-channel interference (CCI) is often caused by a power mismatch

Roaming Design

- Roaming is the method by which client stations move between RF coverage cells in a seamless manner
- Client stations switch communications through different access points
- Seamless communications for client stations moving between the coverage zones is vital for uninterrupted mobility

Roaming Design

- Roaming is one of the most common issues you will need to troubleshoot in WLAN

- Roaming problems are usually caused by poor network design

Roaming Design

- Client stations, not the access point, decide whether or not the client roams between access points

- Some vendors may involve the access point or WLAN controller in the roaming decision, but ultimately the client station initiates the roaming process with a reassociation request frame

Roaming Design

- The method by which a client station decides to roam is a set of proprietary rules determined by the manufacturer of the 802.11 radio, usually defined by a roaming trigger threshold

Roaming Design

- Roaming thresholds usually involve signal strength, SNR, and bit-error rate

- As the client station communicates on the network, it continues to look for other access points via probing and listening on other channels and will hear received signals from other APs

Roaming Design

- The most important variable will always be received signal strength
- As the received signal from the original AP gets weaker and a station hears a stronger signal from another known access point, the station will initiate the roaming process

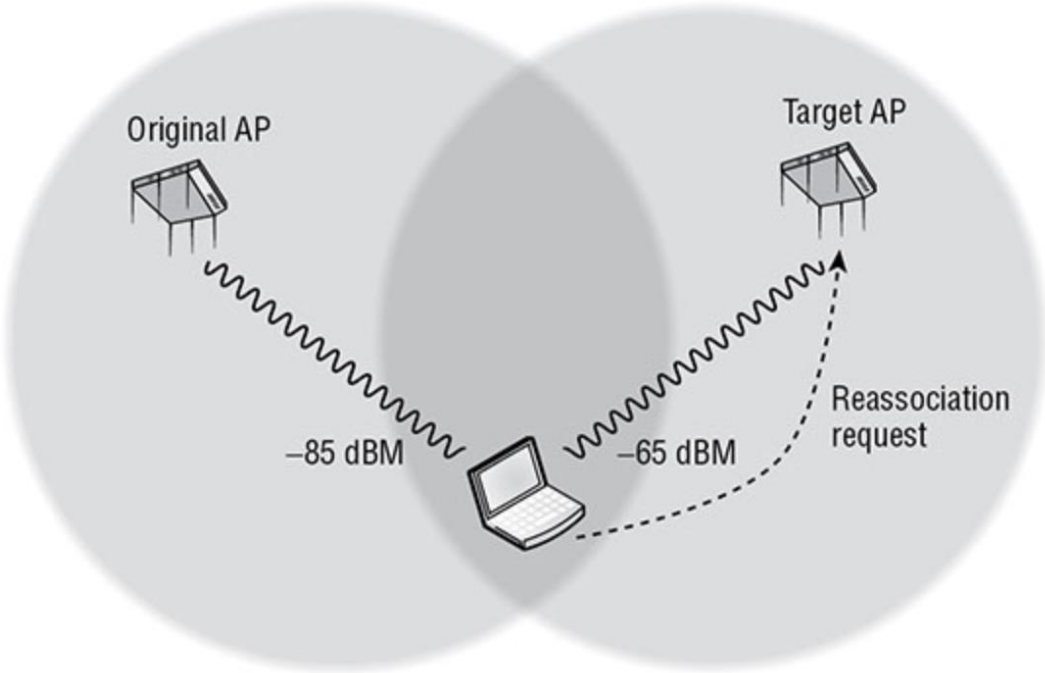
Roaming Design

- However, other metrics, such as SNR, error rates, and retransmissions, may also have a part in what triggers a client to roam

- SNR is a metric used by some WLAN clients to trigger roaming events as well as dynamic rate switching

Roaming Design

- As shown in the right figure, as the client station moves away from the original access point with which it is associated and the signal drops below a predetermined threshold, the client station will attempt to connect to a new target access point that has a stronger signal



Roaming Design – Layer 3 Roaming

- One major consideration when designing a WLAN is what happens when client stations roam across layer 3 boundaries

- Wi-Fi operates at layer 2, and roaming is essentially a layer 2 process

Roaming Design – Layer 3 Roaming

- The roaming is seamless at layer 2, but user VLANs are tied to different subnets on either side of the router

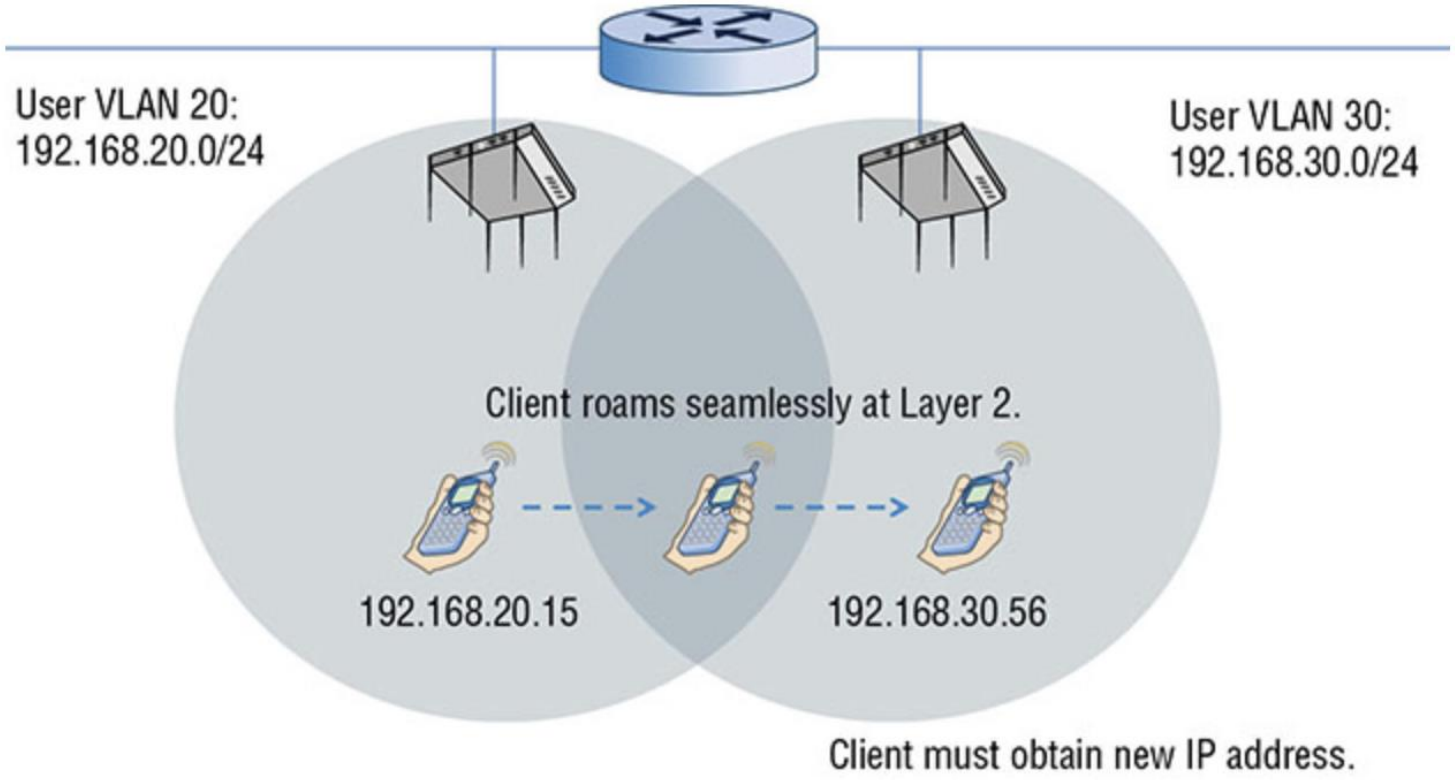
- As a result, the client station will lose layer 3 connectivity and must acquire a new IP address

Roaming Design – Layer 3 Roaming

- Any connection-oriented applications that are running when the client re-establishes layer 3 connectivity will have to be restarted

- For example, a VoIP phone conversation would disconnect in this scenario, and the call would have to be reestablished

Roaming Design – Layer 3 Roaming



Roaming Design – Layer 3 Roaming

- Because 802.11 wireless networks are usually integrated into pre-existing wired topologies, crossing layer 3 boundaries is often a necessity, especially in large deployments

- The only way to maintain upper-layer communications when crossing layer 3 subnets is to provide a layer 3 roaming solution that is based on the Mobile IP standard

Roaming Design – Layer 3 Roaming – Mobile IP

- Mobile IP is an Internet Engineering Task Force (IETF) standard protocol that allows mobile device users to move from one layer 3 network to another while maintaining their original IP address

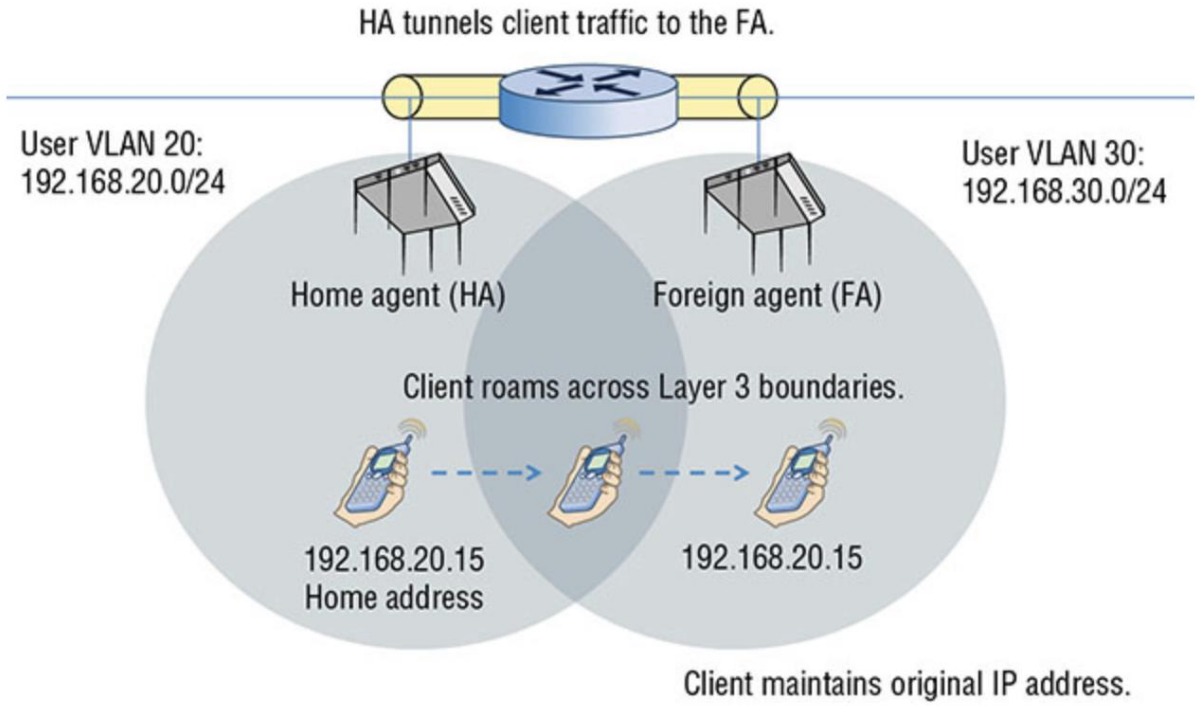
- Mobile IP is defined in RFC 5944

Roaming Design – Layer 3 Roaming – Mobile IP

- Layer 3 roaming solutions based on Mobile IP use some type of tunneling method and IP header encapsulation to allow packets to traverse between separate layer 3 domains, with the goal of maintaining upper-layer communications

Roaming Design – Layer 3 Roaming – Mobile IP

- Most WLAN vendors now support some form of layer 3 roaming solution, as shown in the right figure



Roaming Design – Layer 3 Roaming – Mobile IP

- A mobile client receives an IP address, also known as a home address on a home network

- The mobile client must register its home address with a device called a home agent (HA)

Roaming Design – Layer 3 Roaming – Mobile IP

- Client's original associated access point serves as the home agent
- The home agent is a single point of contact for a client when it roams across layer 3 boundaries
- The HA shares client MAC/IP database information in a table, called a home agent table (HAT) with another device, called the foreign agent (FA)

Roaming Design – Layer 3 Roaming – Mobile IP

- In this example, the foreign agent is another access point that handles all Mobile IP communications with the home agent on behalf of the client
- The foreign agent's IP address is known as the care-of address

Roaming Design – Layer 3 Roaming – Mobile IP

- When the client roams across layer 3 boundaries, the client is roaming to a foreign network where the FA resides
- The FA uses the HAT tables to locate the HA of the mobile client station
- The FA contacts the HA and sets up a Mobile IP tunnel

Roaming Design – Layer 3 Roaming – Mobile IP

- Any traffic that is sent to the client's home address is intercepted by the HA and sent through the Mobile IP tunnel to the FA

- The FA then delivers the tunneled traffic to the client, and the client is able to maintain connectivity using the original home address

Roaming Design – Layer 3 Roaming – Mobile IP

- In our example, the Mobile IP tunnel is between two APs on opposite sides of a router

- If the user VLANs exist at the edge of the network, tunneling of user traffic occurs between access points that assume the roles of HA and FA

Roaming Design – Layer 3 Roaming – Mobile IP

- In a multiple WLAN controller environment, an IP tunnel is created between controllers that are deployed in different routed boundaries with different user VLANs

- One controller functions as the home agent, while another controller functions as the foreign agent

Roaming Design – Layer 3 Roaming

- Although maintaining upper-layer connectivity is possible with these layer 3 roaming solutions, increased latency is sometimes an issue

- Additionally, layer 3 roaming may not be a requirement for your network

Roaming Design – Layer 3 Roaming

- Less complex infrastructure often uses a simpler layer 2 design

- Larger enterprise networks often have multiple user and management VLANs linked to multiple subnets; therefore, a layer 3 roaming solution will be required

Channel Design

- Another key component of WLAN design is the selection of the proper channels to be used among multiple APs in the same location

- A proper channel pattern or channel reuse design is needed to guarantee seamless roaming as well as to prevent two types of interference that are a result of improper channel design

Channel Design - Adjacent Channel Interference

- Adjacent channel interference (ACI) is a degradation of performance resulting from overlapping frequency space that occurs due to an improper channel reuse design
- In the WLAN industry, an adjacent channel is considered to be the next or previous numbered channel
- For example, channel 3 is adjacent to channel 2

Channel Design - Adjacent Channel Interference

- When designing a wireless LAN, you need overlapping coverage cells in order to provide for roaming

- However, the overlapping cells should not have overlapping frequencies, and in the United States only channels 1, 6, and 11 should be used in the 2.4 GHz ISM band to get the most available, non-overlapping channels

Channel Design - Adjacent Channel Interference

- Overlapping coverage cells with overlapping frequencies cause what is known as adjacent channel interference

- If overlapping coverage cells also have frequency overlap from adjacent channels, the transmitted frames will become corrupted, the receivers will not send ACKs, and layer 2 retransmissions will significantly increase

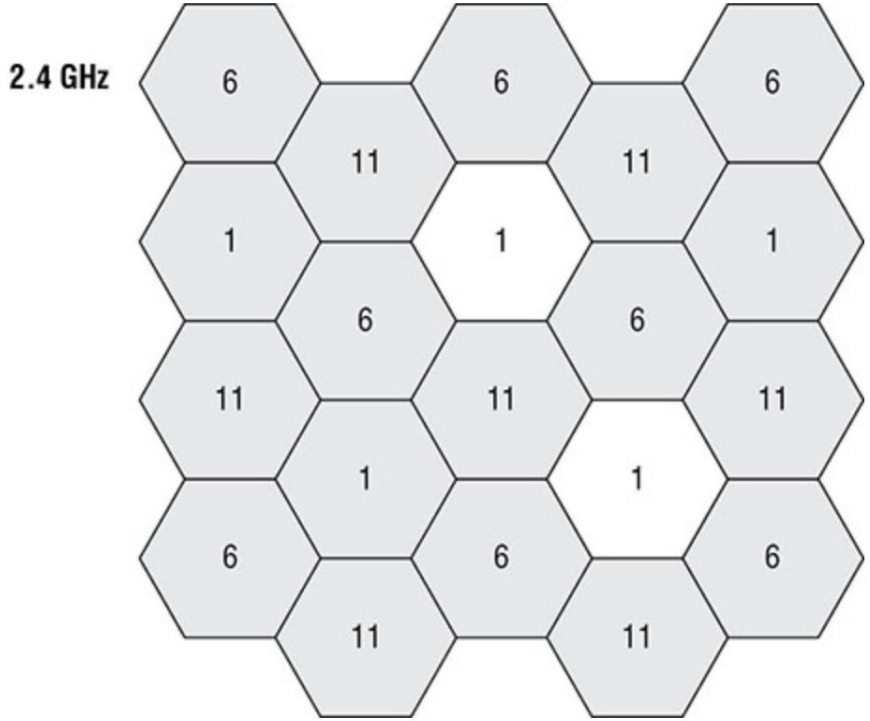
Channel Design - 2.4 GHz Channel Reuse

- To avoid adjacent channel interference, a channel reuse design is necessary
- Once again, overlapping RF coverage cells are needed for roaming, but overlapping frequencies must be avoided
- The only three channels that meet these criteria in the 2.4 GHz ISM band are channels 1, 6, and 11 in the United States

Channel Design - 2.4 GHz Channel Reuse

- APs in the 2.4 GHz band, therefore, should always be placed in a channel reuse pattern similar to the one pictured in the right figure

- Any WLAN channel reuse pattern that uses three or more channels is sometimes referred to as a multiple-channel architecture (MCA)



Channel Design - Co-Channel Interference

- Another of the most common mistakes many businesses make when first deploying a WLAN is to configure multiple access points all on the same channel
- If all the APs are on the same channel, unnecessary medium contention overhead occurs
- CSMA/CA dictates half-duplex communications, and only one radio can transmit on the same channel at any given time

Channel Design - Co-Channel Interference

- If an AP on channel 1 is transmitting, all nearby access points and clients on the same channel within hearing range will defer transmissions

- The result is that throughput is adversely affected:
  - Nearby APs and clients have to wait much longer to transmit because they have to take their turn
  - The unnecessary medium contention overhead that occurs because all the APs are on the same channel is called co-channel interference (CCI)

Channel Design - Co-Channel Interference

- Co-channel interference is almost impossible to prevent in the 2.4 GHz band because only three channels are available for a reuse pattern

- Does an RF signal just stop at the edge of a coverage cell designed for –70 dBm coverage?

  The answer is no, the RF signal continues to propagate, and the signal can be heard by other 802.11 radios at a great distance

Channel Design - Co-Channel Interference

- An 802.11 radio will defer transmissions if it hears the PHY layer preamble transmissions of any other 802.11 radio at a signal detect (SD) threshold just four decibels (dB) or greater above the noise floor

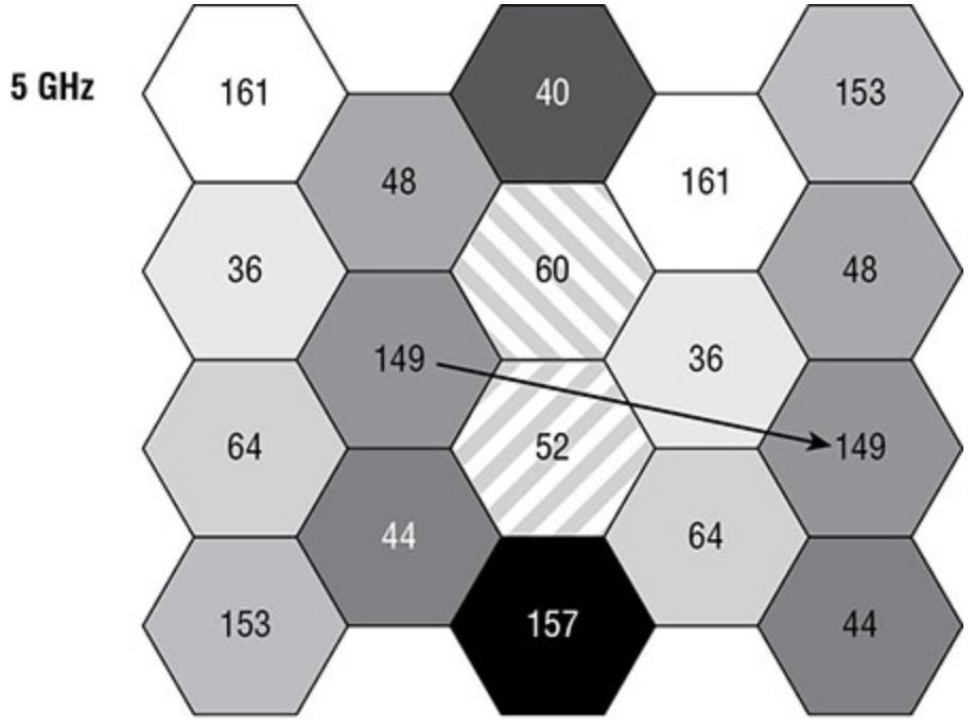
- Any radio that hears another radio on the same channel will defer, which results in medium contention overhead and delay

Channel Design - 5 GHz Channel Reuse

- Channel reuse patterns should also be used in the 5 GHz frequency bands
- If all the 5 GHz channels are legally available for transmissions, a total of 25 channels may be available for a channel reuse pattern at 5 GHz

Channel Design - 5 GHz Channel Reuse

- Depending on the region, and other considerations, 8 channels, 12 channels, 17 channels, 22 channels, or other combinations may be used for 5 GHz channel reuse patterns



Channel Design - 5 GHz Channel Reuse

- Whenever possible, use as many channels as possible in 5 GHz to reduce CCI

- The more channels that are used, the greater the chance that CCI can be prevented, including co-channel interference that originates from client devices

Channel Design - 40 MHz Channel Design

- 802.11n technology introduced the capability of bonding two 20 MHz channels to create a larger 40 MHz channel
- Channel bonding effectively doubles the frequency bandwidth, meaning double the data rates that can be available to 802.11n/ac radios
- Many WLAN access point vendors require that channel bonding be manually enabled because there is the potential for channel bonding to negatively impact the performance of the WLAN

Channel Design - 40 MHz Channel Design

- One of the advantages of using 5 GHz instead of 2.4 GHz is that there are many more 5 GHz 20 MHz channels that can be used in a reuse pattern

- Only three 20 MHz channels can be used in 2.4 GHz. The problem with using only three 20 MHz channels is that there will always be some amount of co-channel interference even though these channels are non-overlapping

Channel Design - 40 MHz Channel Design

- Therefore, a certain amount of medium contention overhead always exists at 2.4 GHz simply because there are not enough channels and frequency space

Channel Design - 40 MHz Channel Design

- Medium contention overhead due to APs on the same 20 MHz channel can be almost completely avoided in 5 GHz because there are more channels

- A 5 GHz channel reuse plan of eight or more 20 MHz channels will greatly decrease co-channel interference and medium contention overhead

Channel Design - 40 MHz Channel Design

- If only eight 20 MHz channels are being used, then a four channel 40 MHz channel reuse pattern exists in 5GHz

Channel Design - 40 MHz Channel Design

- Although the bandwidth is doubled for the 802.11n/ac radios, there will be an increase of medium contention overhead because there are only four 40 MHz channels, and access points and clients on the same 40 MHz channel will likely hear each other

- The medium contention overhead may have a negative impact and decreases any gains in performance that the extra bandwidth might provide

Channel Design - 40 MHz Channel Design

- Another problem with channel bonding is that it usually will result in a higher noise floor of about 3 dB
- If the noise floor is 3 dB higher, then the SNR is 3 dB lower, which means that the radios may shift down to lower MCS rates and therefore lower modulation data rates
- In many cases, this offsets some of the bandwidth gains that the 40 MHz frequency space provides

Channel Design - 40 MHz Channel Design

- So, should you use channel bonding or not? If four or fewer 40 MHz channels are available, you might not want to turn on channel bonding, especially if the 5 GHz radios are transmitting at a higher power level

- If the majority of the WLAN clients do not support channel bonding, don’t use channel bonding

Channel Design - Static Channels and Transmit Power vs. Adaptive RF (RRM)

- Probably the most debated topic when it comes to WLAN design is whether to use static channel and power settings for APs versus using adaptive channel and power settings

- Radio resource management (RRM) is an industry standard term used to describe the automatic and adaptive power and channel configuration of access points

Channel Design - Static Channels and Transmit Power vs. Adaptive RF (RRM)

- WLAN vendor APs can dynamically change their configuration based on accumulated RF information gathered from the access point radios

- Based on the accumulated RF information, the access points adjust their power and channel settings, adaptively changing the RF coverage cells

Channel Design - Static Channels and Transmit Power vs. Adaptive RF (RRM)

- Radio resource management is also referred to as adaptive RF

- When implemented, RRM provides automatic cell sizing and automatic monitoring and optimization of the RF environment, which can best be described as a self-organizing wireless LAN

Channel Design - Static Channels and Transmit Power vs. Adaptive RF (RRM)

- Adaptive RF capabilities are turned on by default on most WLAN vendor AP

- The algorithms for RRM constantly improve year after year. The majority of commercial WLAN customers use RRM because it is easy to deploy

Channel Design - Static Channels and Transmit Power vs. Adaptive RF (RRM)

- RRM is usually the preferred method in enterprise deployments with thousands of APs

- However, careful consideration should be given to using static channel and power settings in complex RF environments

Channel Design – Single Channel Architecture

- In Single Channel Architecture, the client stations see transmissions on only a single channel with one SSID (logical WLAN identifier) and one BSSID (layer 2 identifier)

- From the perspective of the client station, only one access point exists

Channel Design - Single-Channel Architecture

- In this type of WLAN architecture, all access points in the network can be deployed on one channel in 2.4 GHz or 5 GHz frequency bands

Channel Design - Single-Channel Architecture

- Uplink and downlink transmissions are coordinated by a WLAN controller on a single 802.11 channel in such a manner that the effects of co-channel interference are minimized

Channel Design - Single-Channel Architecture

- Client stations believe they are connected to only a single access point, although they may be roaming across multiple physical Access Points

Channel Design - Single-Channel Architecture

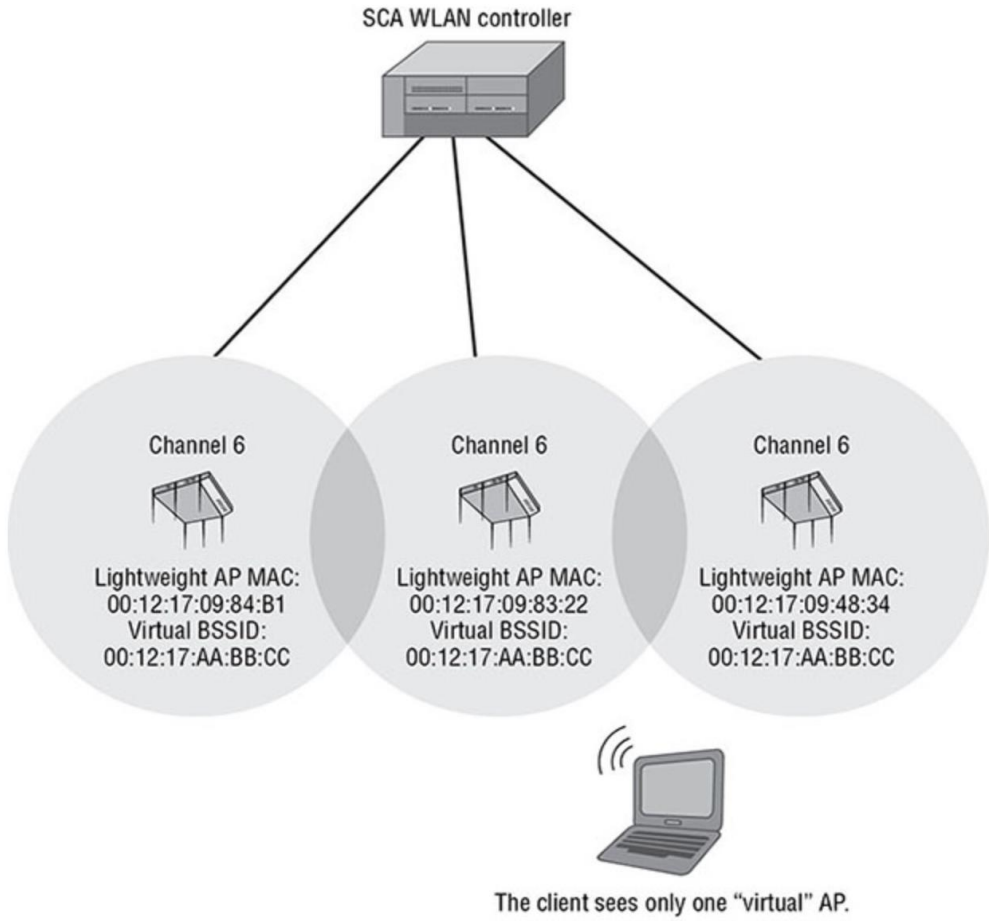
- As earlier said, clients make the roaming decisions

- In a single-channel architecture (SCA) system, the clients think they are associated to only one AP, so they never initiate a layer 2 roaming exchange. All the roaming handoffs are handled by a central WLAN controller

Channel Design - Single-Channel Architecture

- The main advantage is that clients experience zero handoff time, and the latency issues associated with roaming times are resolved

- The virtual AP used by SCA solutions is potentially an excellent marriage for VoWiFi phones



Channel Design - Single-Channel Architecture

- We know that client stations make the roaming decision in an MCA environment

- However, client stations do not know that they roam in an SCA environment

Channel Design - Single-Channel Architecture

- The clients must still be mobile and transfer layer 2 communications between physical access points

- All the client-roaming mechanisms are now handled back on the WLAN controller, and client-side roaming decisions have been eliminated in Single Channel Architecture

Capacity Design

- When a wireless network is designed, two concepts that typically compete with each other: **Capacity and Range**

Capacity Design

- In the early days of wireless networks, it was common to install an access point with the power set to the maximum level to provide the largest coverage area possible

Capacity Design

- This was typically acceptable because there were very few wireless devices
- Also, access points were very expensive, so companies tried to provide the most coverage while using the fewest access points

Capacity Design

- Most WLANs are designed with a primary focus on client capacity needs
- This does not mean that coverage design is now ignored
- We still plan for a –70 dBm or better received signal, high SNR, seamless roaming, and a proper channel reuse pattern

Capacity Design

- Important to know that, how you design for coverage will also impact capacity needs

- As it is said before, APs configured for full transmit power are no longer ideal

Capacity Design

- Adjusting the AP transmit power to limit the effective coverage area is known as cell sizing and is one of the most common methods of meeting client capacity needs

- Typical indoor WLAN deployments are designed with the APs set at about one-fourth to one-third transmit power

Capacity Design

- Higher user and client density environments may require that the AP transmit power be set at the lowest setting of 1 mW

- In other words, more APs are needed to meet capacity needs, and therefore AP transmit power will need to be lowered

Capacity Design

- Limiting the transmit power of APs also helps reduce CCI caused by APs, which has a direct impact on performance

Capacity Design - High Density

- The terms high density (HD) and very high density (VHD) are often used when discussing capacity design and planning for a WLAN

Capacity Design - High Density

- Different WLAN engineers have different opinions as to what constitutes a high-density WLAN, however, due to the huge number of client devices, most WLANs should be considered high density by default

Capacity Design - High Density

- The average person may want to connect to an enterprise WLAN with as many as three or four Wi-Fi devices

- Obviously, the density of client devices also depends on the number of users

Capacity Design - High Density

- Most high-density environments consist of multiple areas where roaming is also a top priority

- APs are deployed in many different rooms with walls that will often contribute to different levels of attenuation

Capacity Design - Very High Density

- Any WLAN environment that has a tremendous amount of people in a single open area is often referred to as a very high-density (VHD) WLAN

- Prime examples include auditoriums, gymnasiums, cafeterias, etc.

Capacity Design - Very High Density

- Most VHD environments do not have walls that provide attenuation
- All the APs likely hear each other within the open space
- Design for a very high-density WLAN is quite complex and different from standard high-density environments with walls

Capacity Design - Ultra High Density

- An ultra high-density WLAN is defined as an environment with tens of thousands of users and devices all within the same space

- The best examples of an ultra high-density WLAN are stadiums and sports arenas

Capacity Design - How many clients can connect to an AP?

- There are simply too many variables to always give the same answer for any WLAN vendor's AP

- The default settings of an enterprise WLAN radio might allow as many as 100–250 client connections

Capacity Design - How many clients can connect to an AP?

- Since most enterprise APs are dual-frequency with both a 2.4 GHz and 5 GHz radio, theoretically 200–500 clients could associate to the radios of a single AP

Capacity Design - How many clients can connect to an AP?

- Although more than a hundred devices might be able to connect to an AP radio, these numbers are not realistic for active devices due to the nature of the half-duplex shared medium

Capacity Design - How many clients can connect to an AP?

- The performance needs of this many client devices will not be met and the user experience will be miserable

- The perception will be that the Wi-Fi is “slow” If the access point is using 802.11n/ac radios with 20 MHz channels, a good rule of thumb is that each radio could support 35–50 active devices for average use

Capacity Design – Common Design Questions

- Below questions should be asked while designing WLAN for capacity

1. What type of applications will be used on the WLAN?
2. How many users and devices are expected?
3. What types of client devices are connecting to the WLAN?

Capacity Design - What type of applications will be used on the WLAN?

- As previously stated, 35–50 active Wi-Fi devices per radio, communicating through a dual-frequency 802.11n/ac access point, with average application use, such as web browsing and email, is realistic

Application	Required Throughput
Email/web browsing	500 Kbps to 1 Mbps
Printing	1 Mbps
SD video streaming	1 Mbps to 1.5 Mbps
HD video streaming	2 Mbps to 5 Mbps

Capacity Design - What type of applications will be used on the WLAN?

- However, bandwidth-intensive applications, such as high-definition video streaming, will have an impact
- Different applications require different amounts of throughput

Application	Required Throughput
Email/web browsing	500 Kbps to 1 Mbps
Printing	1 Mbps
SD video streaming	1 Mbps to 1.5 Mbps
HD video streaming	2 Mbps to 5 Mbps

Capacity Design - How many users and devices are expected?

Three important questions need to be asked with regard to users

1. First, how many users currently need wireless access and how many Wi-Fi devices will they be using?
2. Second, how many users and devices may need wireless access in the future?
3. Where are the users?

These first two questions will help you to begin adequately planning for a good ratio of devices per access point while allowing for future growth

Capacity Design - How many users and devices are expected?

- The third question of great significance is, where are the users? Sit down with network management and indicate on the floor plan of the building any areas of high user density

- For example, one company might have offices with only 1 or 2 people per room, whereas another company might have 30 or more people in a common area separated by cubicle walls

Capacity Design - How many users and devices are expected?

- Other examples of areas with high user density are call centers, classrooms, and lecture halls

Capacity Design - How many users and devices are expected?

- You should always plan to conduct a validation survey when the users are present, not during off-hours

Capacity Design - How many users and devices are expected?

- A high concentration of human bodies can attenuate the RF signal because of absorption

Capacity Design - What types of client devices are connecting to the WLAN?

- Always remember that all client devices are not equal. Many client devices consume more airtime due to lesser MIMO capabilities

- For example, an older 802.11n tablet with a 1×1:1 MIMO radio transmitting on a 20 MHz channel can achieve a data rate of 65 Mbps with TCP throughput of 30 Mbps to 40 Mbps

Capacity Design - What types of client devices are connecting to the WLAN?

- An 802.11n tablet with a 2×2:2 MIMO radio transmitting on a 20 MHz channel might achieve a data rate of 130 Mbps with TCP throughput of 60 Mbps to 70 Mbps
- Many laptops also have 3×3:3 MIMO capabilities and thus are capable of higher data rates
- The bulk of newer smartphones and tablets are now 2×2:2 MIMO capable

Capacity Design - What types of client devices are connecting to the WLAN?

- The point is that devices with less MIMO capabilities consume more airtime and therefore affect the aggregate performance of any WLAN

- An AP can service more 2×2:2 MIMO clients efficiently as opposed to legacy 1×1:1 MIMO clients, which operate at lower data rates

Capacity Design – How many AP per room?

- The location where APs are physically mounted also needs to be taken into consideration based on capacity needs
- Some large areas of a building may only have two or three users that require Wi-Fi access
- Conversely, other areas, such as an auditorium may have hundreds of users and devices that need Wi-Fi connectivity

Capacity Design – How many AP per room?

- In education environment, due to capacity requirements, it has become commonplace to deploy one AP per room
- Please note that one AP per classroom maybe entirely unnecessary
- One AP per every two or three classrooms may be sufficient to meet capacity needs

Capacity Design – How many AP per room?

- How many APs are needed depends on the capacity requirements as well as customer environment

- Do you need to deploy one AP in every room?

  It depends on the number of devices, the type of devices, and the application traffic

Capacity Design - Band Steering

- The unlicensed 5 GHz frequency spectrum offers many advantages over the unlicensed 2.4 GHz frequency spectrum for Wi-Fi communications

- The 5 GHz U-NII bands offer a wider range of frequency space and many more channels

Capacity Design - Band Steering

- A proper 5 GHz channel reuse pattern using multiple channels will greatly decrease medium contention overhead caused by co-channel interference

- In the 2.4 GHz band, there is always medium contention overhead due to CCI simply because there are only three channels

Capacity Design - Band Steering

- Another consideration of the 2.4 GHz band is that, in addition to this band being used for WLAN networking, it is heavily used by many other types of devices, including microwave ovens, baby monitors, cordless telephones, and video cameras

Capacity Design - Band Steering

- With all of these different devices operating in the same frequency range, there is much more RF interference and a much higher noise floor than in the 5 GHz bands

Capacity Design - Band Steering

- So, if the use of the 5 GHz bands will provide better throughput and performance, how can we encourage the clients to use this band?

- For starters, it is the client that decides which AP and which band to connect to, typically based on the strongest signal that it hears for the SSID that it wants to connect to

Capacity Design - Band Steering

- Most access points have both 2.4 GHz and 5 GHz radios in them, with both of them advertising the same SSIDs

- Since the 5 GHz signals naturally attenuate more than the 2.4 GHz signals, it is likely that the client radio will identify the 2.4 GHz radio as having a stronger signal and connect to it by default

Capacity Design - Band Steering

- In many environments, the client would be capable of making a strong and fast connection with either of the AP's radios but will choose the 2.4 GHz signal because it is the strongest

Capacity Design - Band Steering

- A technology known as band steering has been developed to try to encourage dual-band client radios to connect to a 5 GHz AP radio instead of a 2.4 GHz AP radio

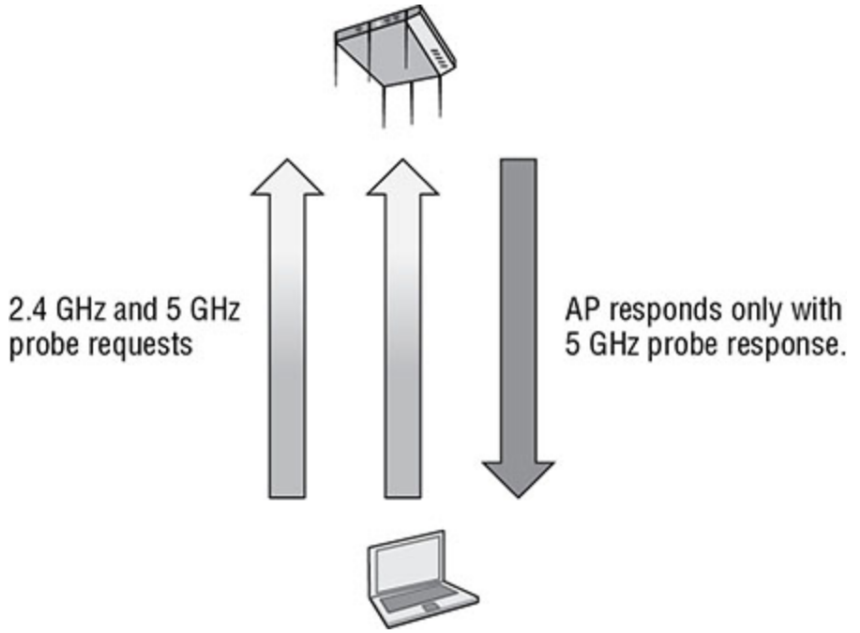
Capacity Design – How Band Steering works?

- When a dual-frequency client first starts up, it will transmit probe requests on both the 2.4 and 5 GHz bands, looking for an AP

- When a dual-frequency AP hears probe requests on both bands originating from the same client, the AP knows that the client is capable of operating in the 5 GHz band

Capacity Design – How Band Steering works?

- As depicted in the right figure, the AP will then try to steer the client to the 5 GHz band by responding to the client using only 5 GHz transmissions



Capacity Design – Load Balancing

- Load balancing the clients between access points, ensures that a single AP is not overloaded with too many clients, and that the total client population can be served by numerous APs, with the final result being better performance

Capacity Design – Load Balancing

- When a client wants to connect to an AP, the client will send an association request frame to the AP
- If an AP is already overloaded with too many clients, the AP will ignore the association response of the client
- Assumption is that the client will then send another association request to another nearby AP with a lesser client load
- Over time, the client associations will be fairly balanced across multiple APs

Capacity Design – Load Balancing

- The client load information will obviously have to be shared among the access points

- Load balancing is a control plane mechanism that can exist either in a distributed architecture where all the APs communicate with each other, or within a centralized architecture that utilizes a WLAN controller

Capacity Design - Airtime Consumption

- Designing for client device capacity has now become the norm

- WLAN design practices now dictate that you design to minimize airtime consumption, which is directly related to capacity design

Capacity Design - Airtime Consumption

- Wi-Fi is a half-duplex RF medium and only one radio can transmit on a channel at any given time

Capacity Design - Airtime Consumption

- Whenever a radio has won a transmission opportunity, the radio owns the available airtime until it finishes transmitting

- Yes, every radio needs to be able to transmit and deliver data; however, there are some simple WLAN design best practices that can minimize unnecessary airtime consumption

Capacity Design - Airtime Consumption

- Co-channel interference (CCI) is the top cause of unnecessary airtime consumption that can be minimized with proper WLAN design best practices

- Designing for –70 dBm coverage and high SNR also ensures that client devices will transmit 802.11 data frames at high data rates, based on the client radio's capabilities

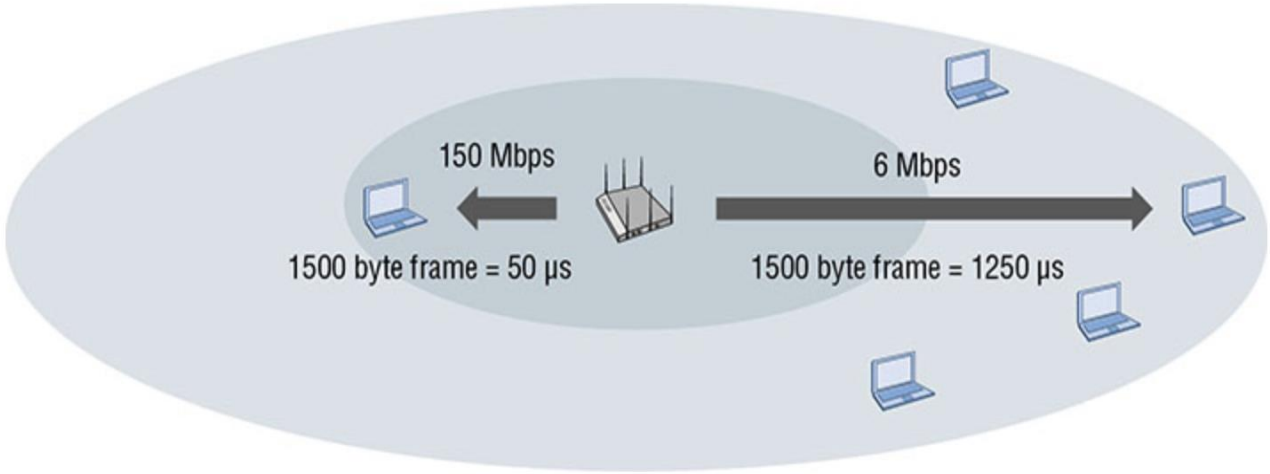
Capacity Design - Airtime Consumption

- So what are some other WLAN design best practices that can reduce airtime consumption?

- One of the best ways to cut down on airtime consumption is to disable some of the lower data rates on an AP

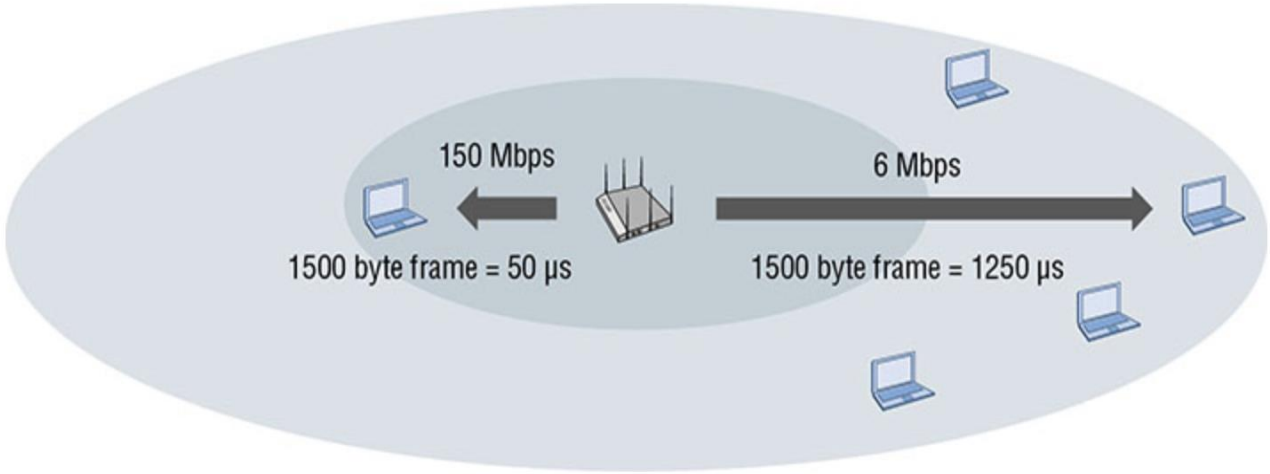
Capacity Design - Airtime Consumption

- Figure depicts an AP communicating with multiple client stations at 6 Mbps while communicating with one client using a 150 Mbps data rate



Capacity Design - Airtime Consumption

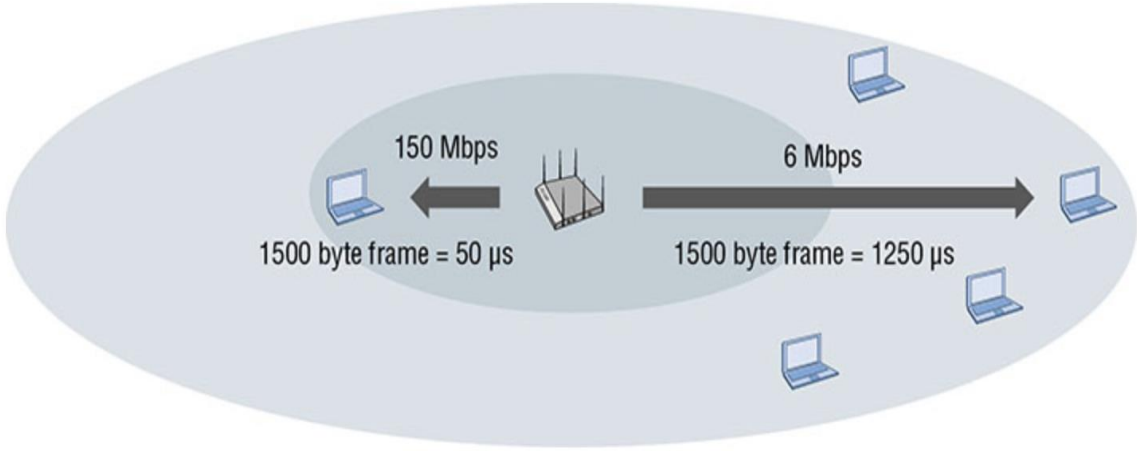
- When 802.11 radios transmit at very low data rates, such as 6 Mbps or even lower, they effectively cause medium contention overhead for higher data rate transmitters due to the long wait time



Capacity Design - Airtime Consumption

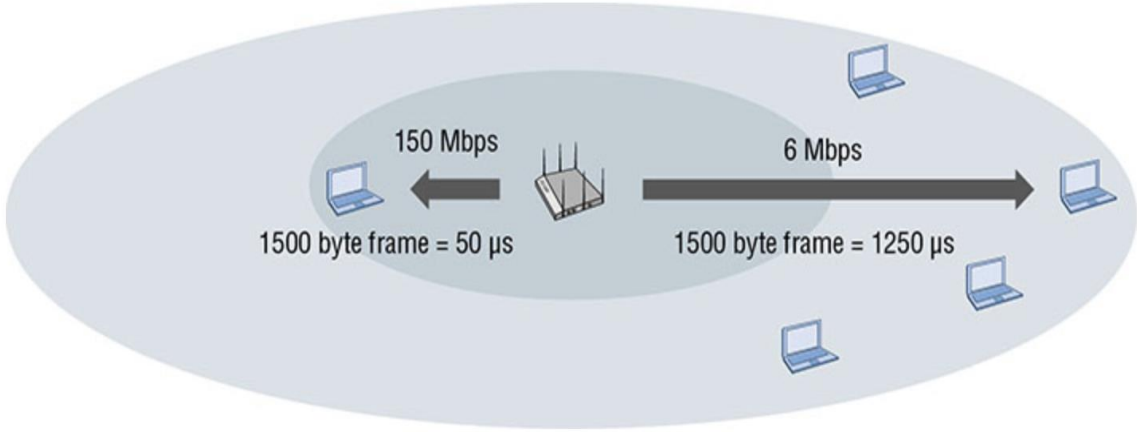
- A radio transmitting a 1,500-byte data frame at 150 Mbps might occupy the medium for 50 microseconds

- Another radio transmitting at 6 Mbps may take 1,250 microseconds to deliver the same 1,500 bytes



Capacity Design - Airtime Consumption

- In other words, the same data payload consumes 2,500 percent more airtime when being delivered at the lower data rate



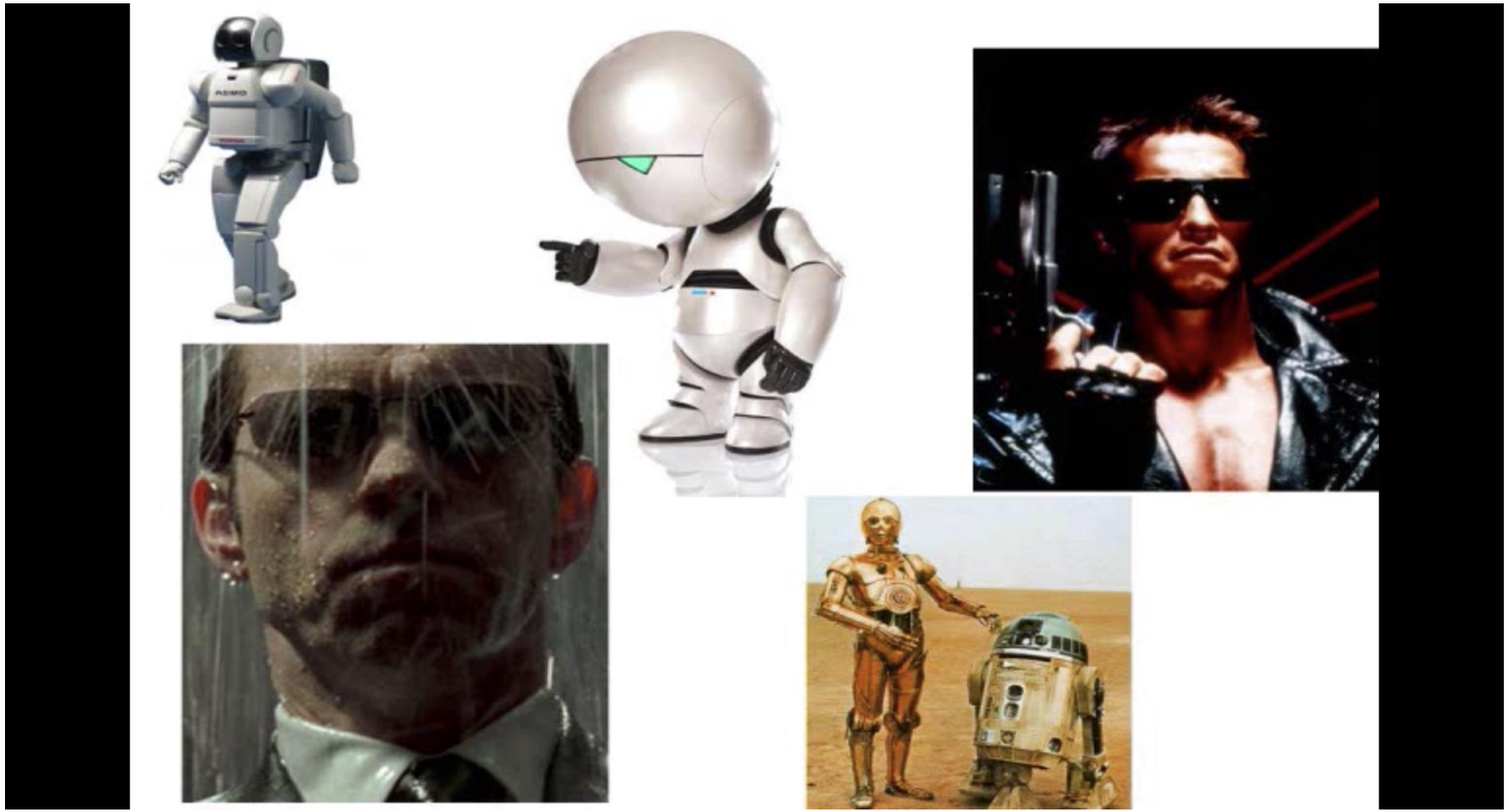
Artificial Intelligence Machine Learning and Deep Learning

What is AI , ML and DL

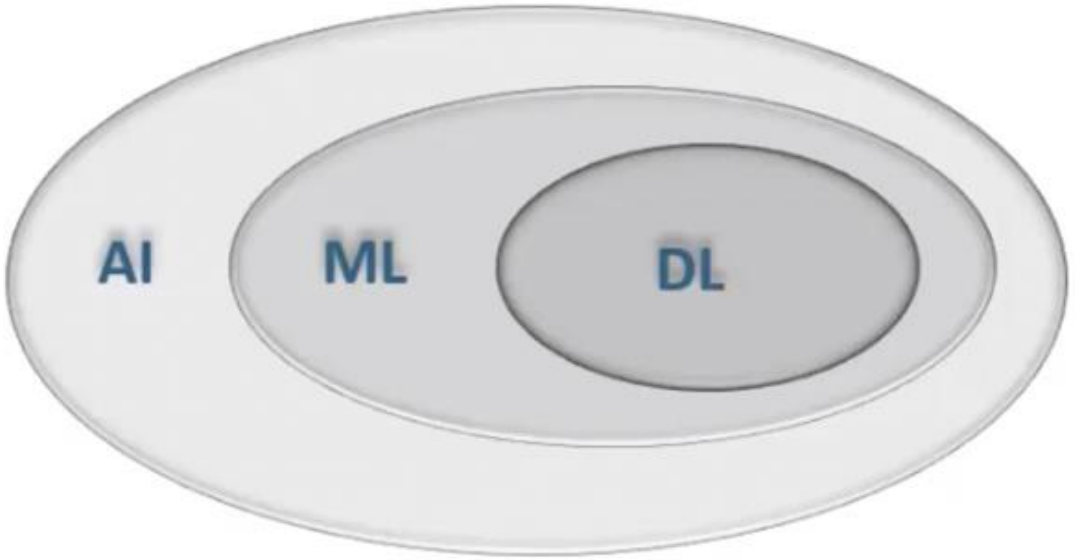
- Artificial intelligence (AI) makes it possible for machines to learn from experience, adjust to new inputs and perform human-like tasks

- Machine Learning and Deep Learning are subset of Artificial Intelligence

Artificial Intelligence



What is AI , ML and DL



ML and DL are subsets of AI

You perform AI using ML and DL technology

Artificial Intelligence Applications

Voice Recognition

NETFLIX

Content Recommendation

Google Translate

Autonomous Driving

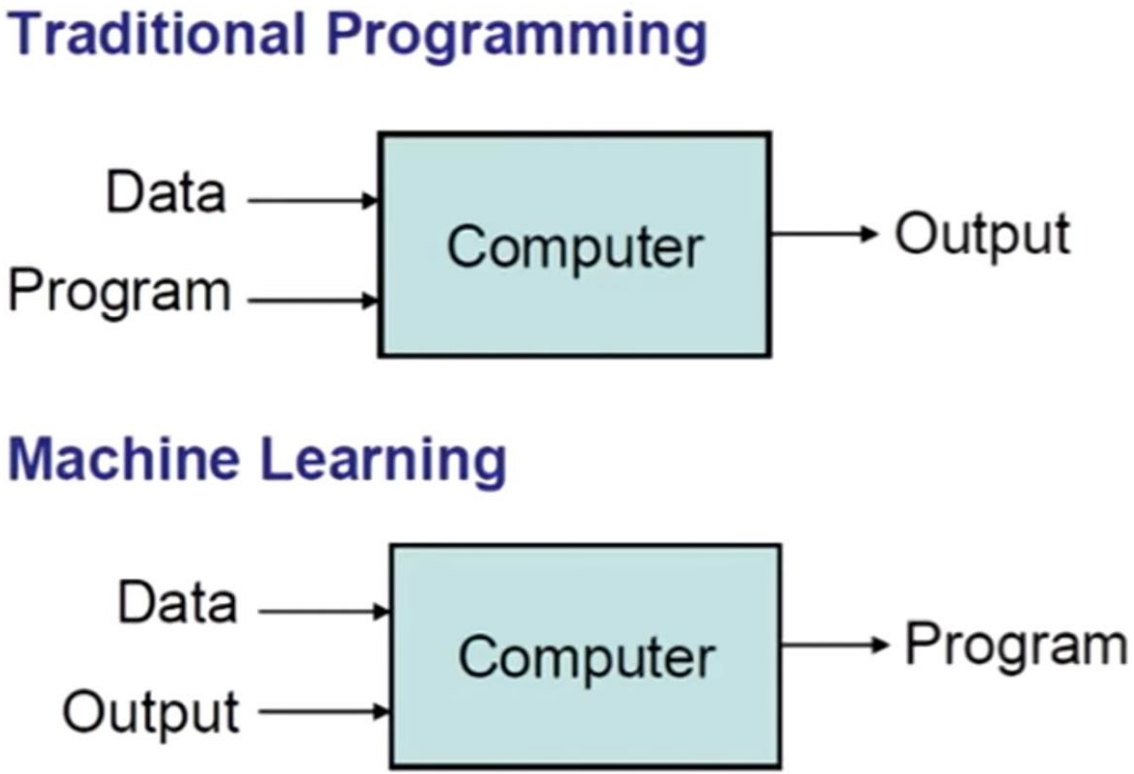
Drivers of Advances in AI

Processing

Algorithms

A Lot of Data

Machine Learning vs. Traditional Programming

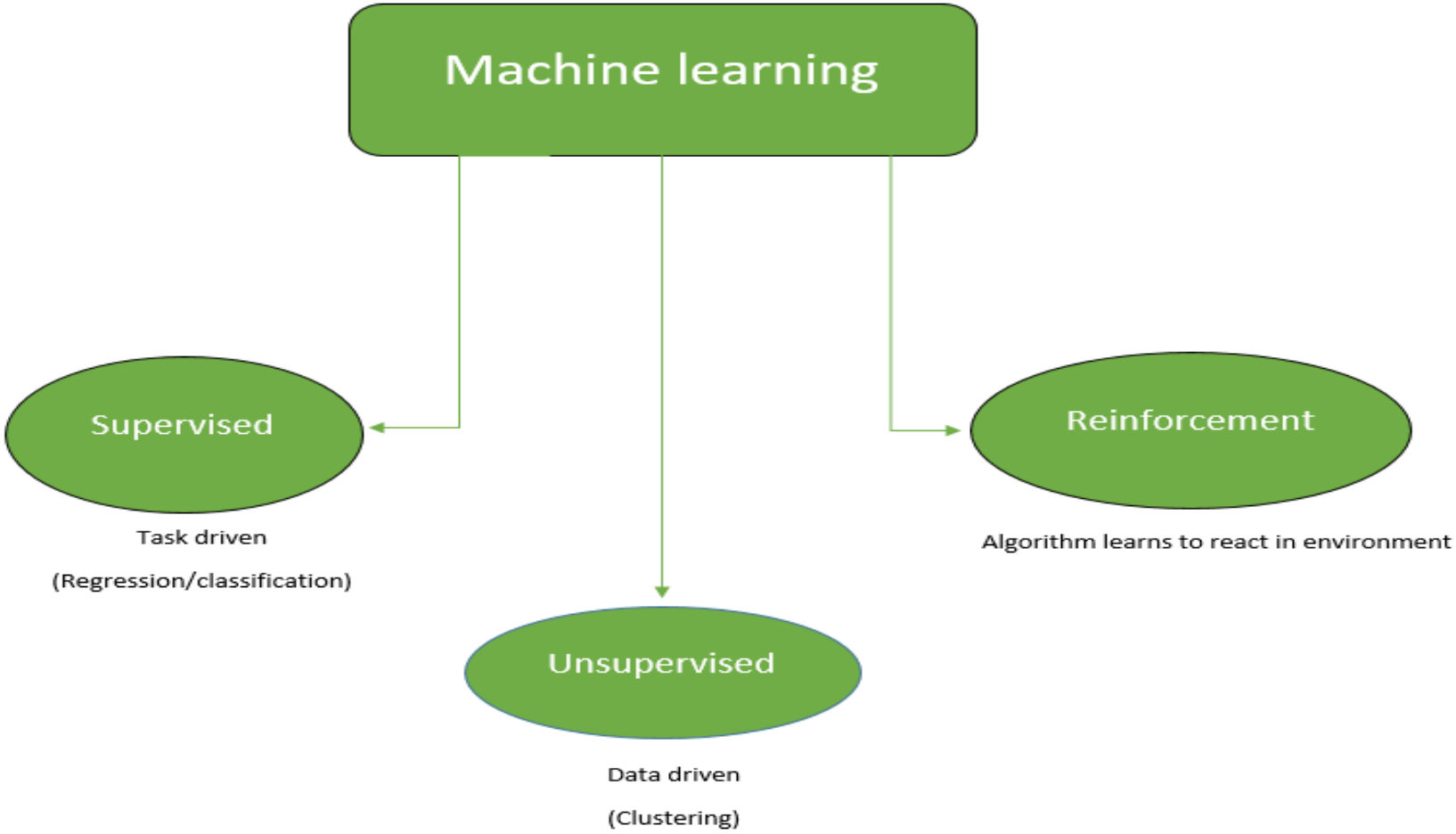


Machine Learning

- Machine Learning is an algorithm that learn from data without explicitly programmed

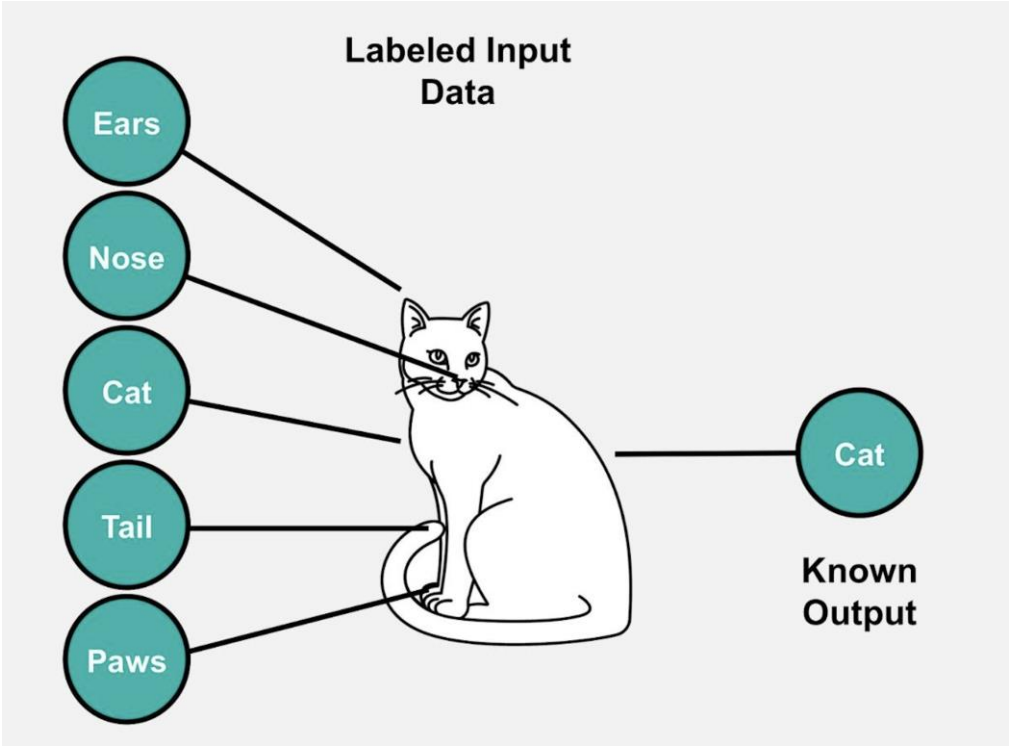
- There are many types of Machine Learning but mainly all Machine Learning and Deep Learning algorithms are classified as either Supervised or Unsupervised Learning or Reinforcement Learning

Machine Learning Types



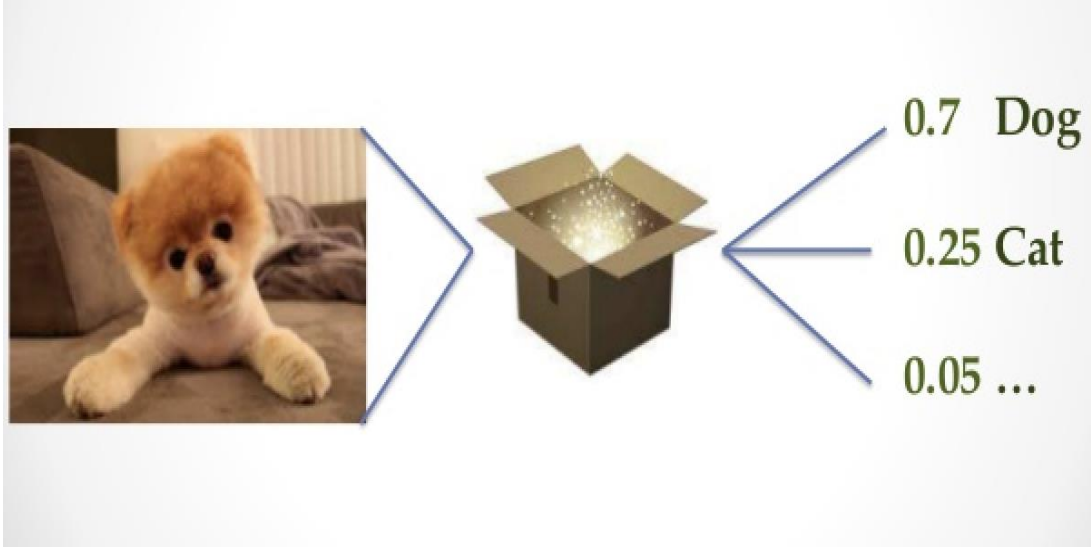
Machine Learning – Supervised Learning

Supervised learning uses labeled training data to train machines to learn relationships between given inputs to a given output



Machine Learning – Supervised Learning

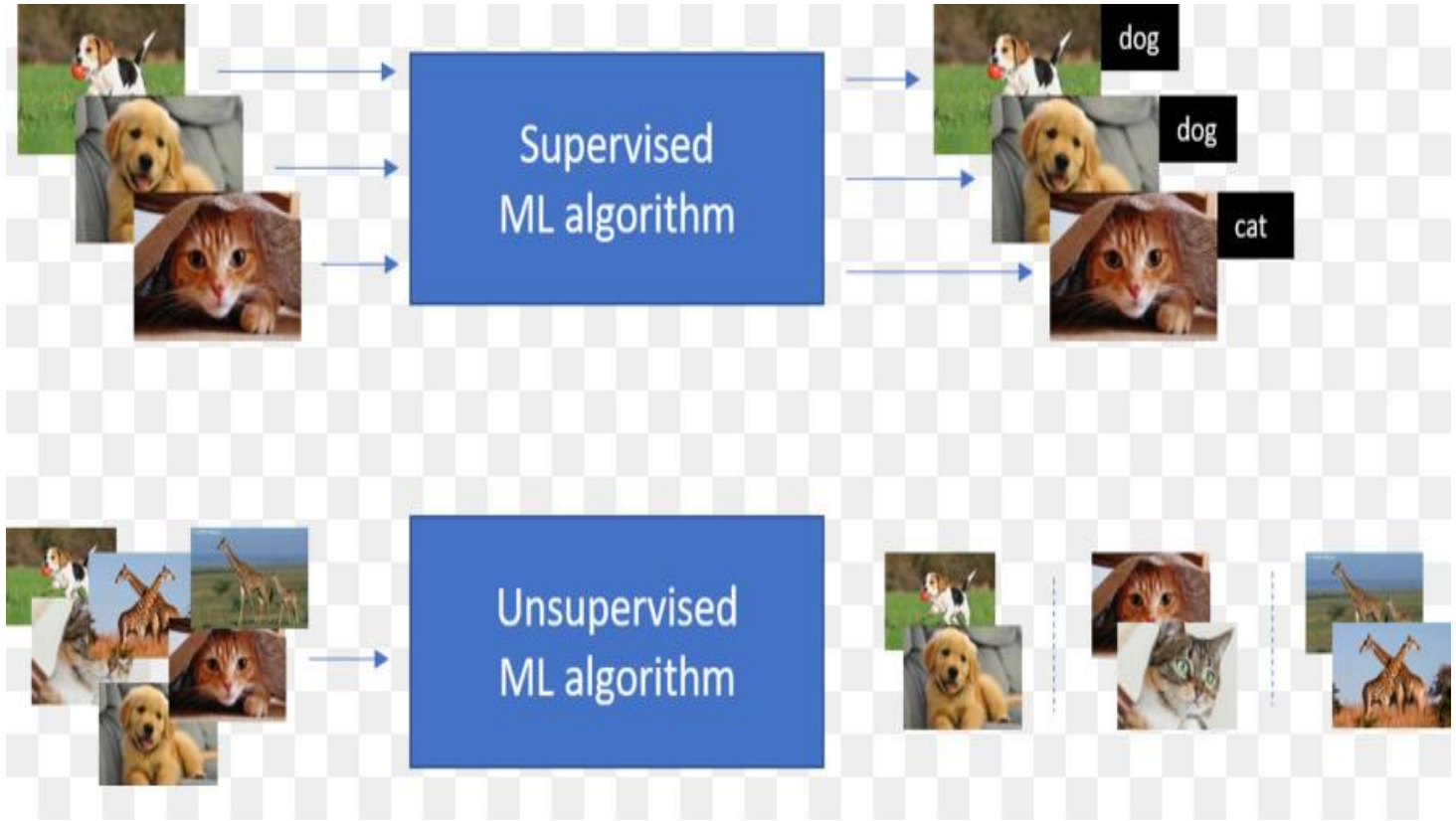
Supervised Learning provides predictability for each labeled information



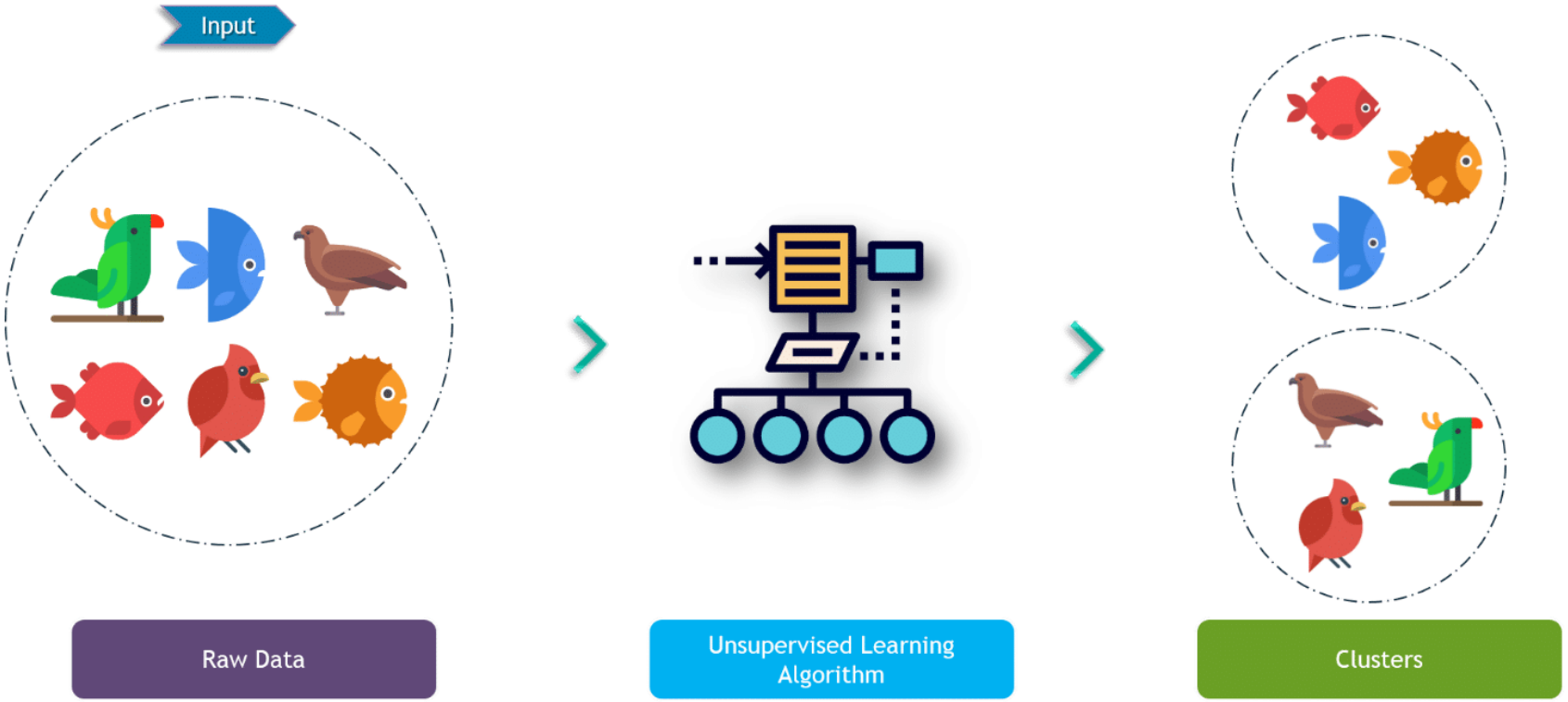
Machine Learning – Supervised Learning

Supervised vs. Unsupervised Learning

- Supervised Learning separate each labeled data
- Unsupervised Learning cluster them



Machine Learning – Unsupervised Learning



Machine Learning – Unsupervised Learning

Machine Learning – Reinforcement Learning

- Reinforcement learning (RL) is learning by interacting with an environment
- An RL agent learns from the consequences of its actions and it selects its actions on basis of its past experiences and also by new choices (exploration)
- Reinforcement Learning is essentially trial and error learning

Machine Learning – Reinforcement Learning

- The reinforcement signal that the RL-agent receives is a numerical reward, which encodes the success of an action's outcome, and the agent seeks to learn to select actions that maximize the accumulated reward over time

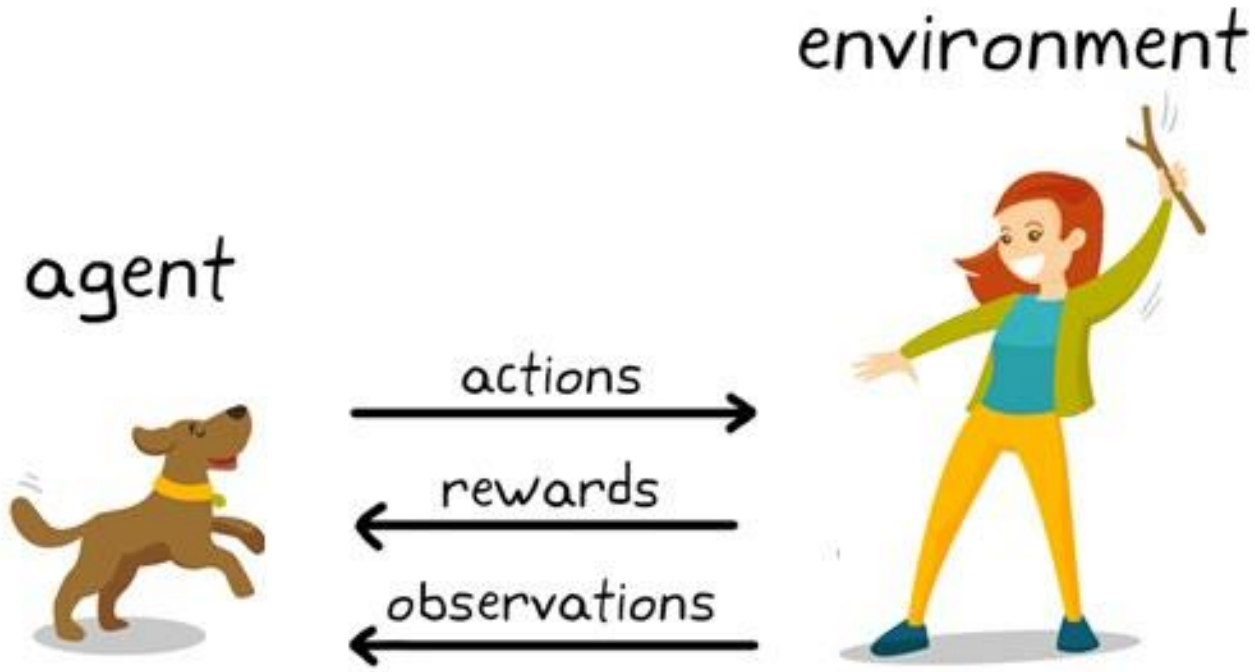
Machine Learning – Reinforcement Learning

- Dog training can be given as an example to Reinforcement Learning

- Whole meaning of reinforcement learning training is to “tune” the dog’s action so that it learns the desired behaviors by getting a reward when each time it performs correct action

Machine Learning – Reinforcement Learning

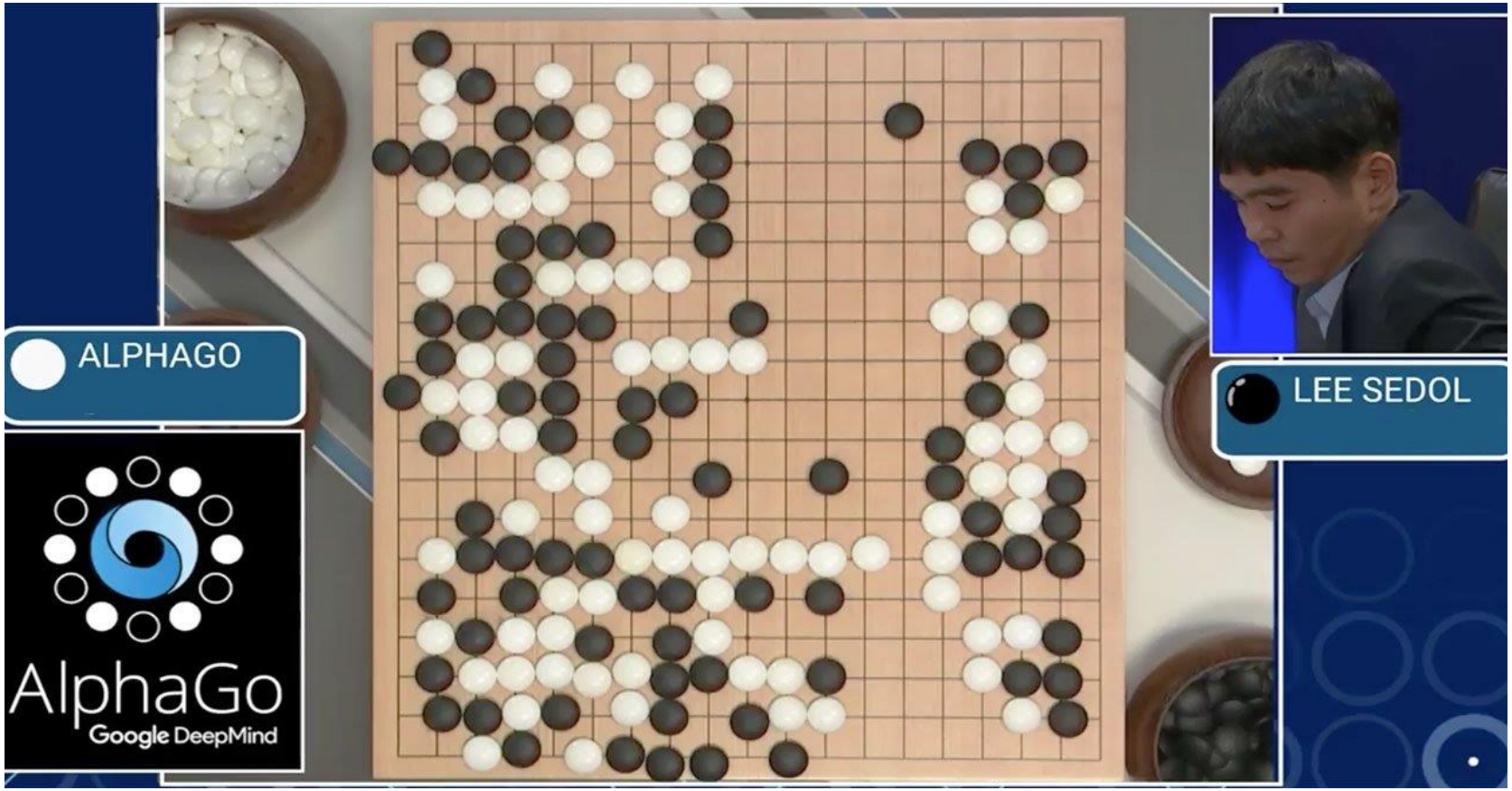
- After training is complete, the dog should be able to observe the owner and take the appropriate action, for example, sitting when command is to “sit”



Real Life example of Reinforcement Learning

Autonomous Parking

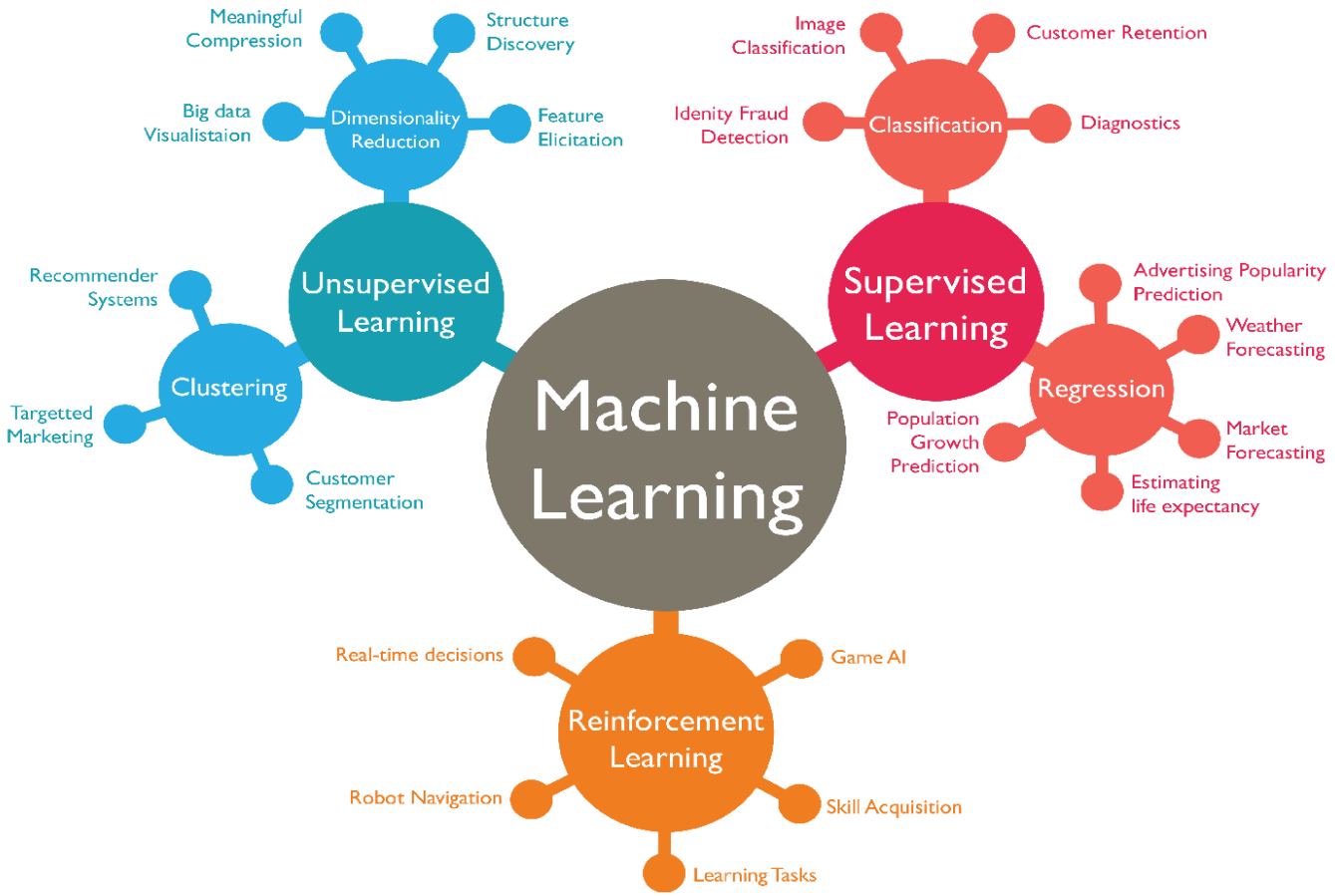
Real Life example of Reinforcement Learning



Real Life Examples of Supervised and Unsupervised Learning

Supervised and Unsupervised Learning are not just used to understand whether picture is cat or dog!

There are many real life applications which people can get benefit!

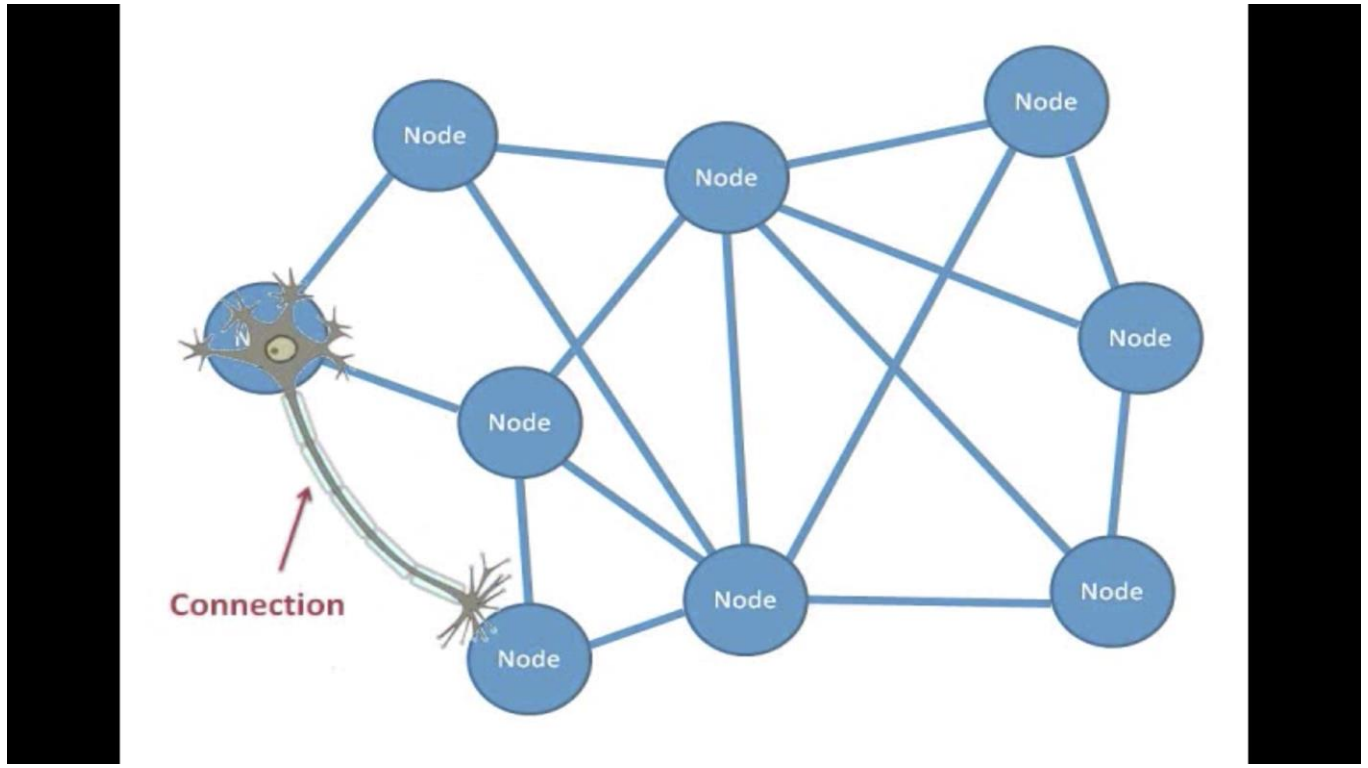


Deep Learning (Neural Networks)

- Deep learning is a machine learning technique
- It teaches a computer to filter inputs through layers to learn how to predict and classify information
- Its purpose is to mimic how the human brain works, thus commonly knowns as Deep Neural Networks

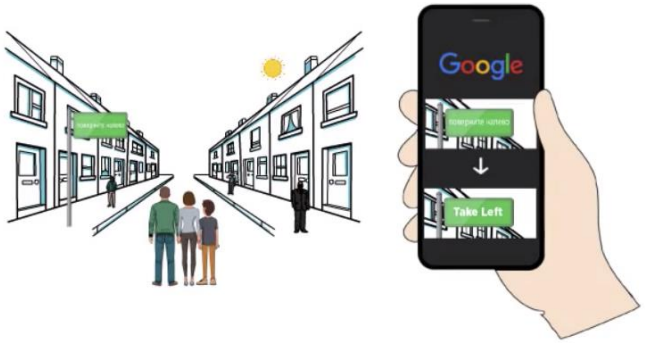
Deep Learning (Neural Networks)

- In the human brain, there are about 100 billion neurons. Each neuron connects to about 100.000 of its neighbors. We are recreating that, but in a way and at a level that works for machines



Deep Learning (Neural Networks) Applications

Facial recognition



Real-time translation

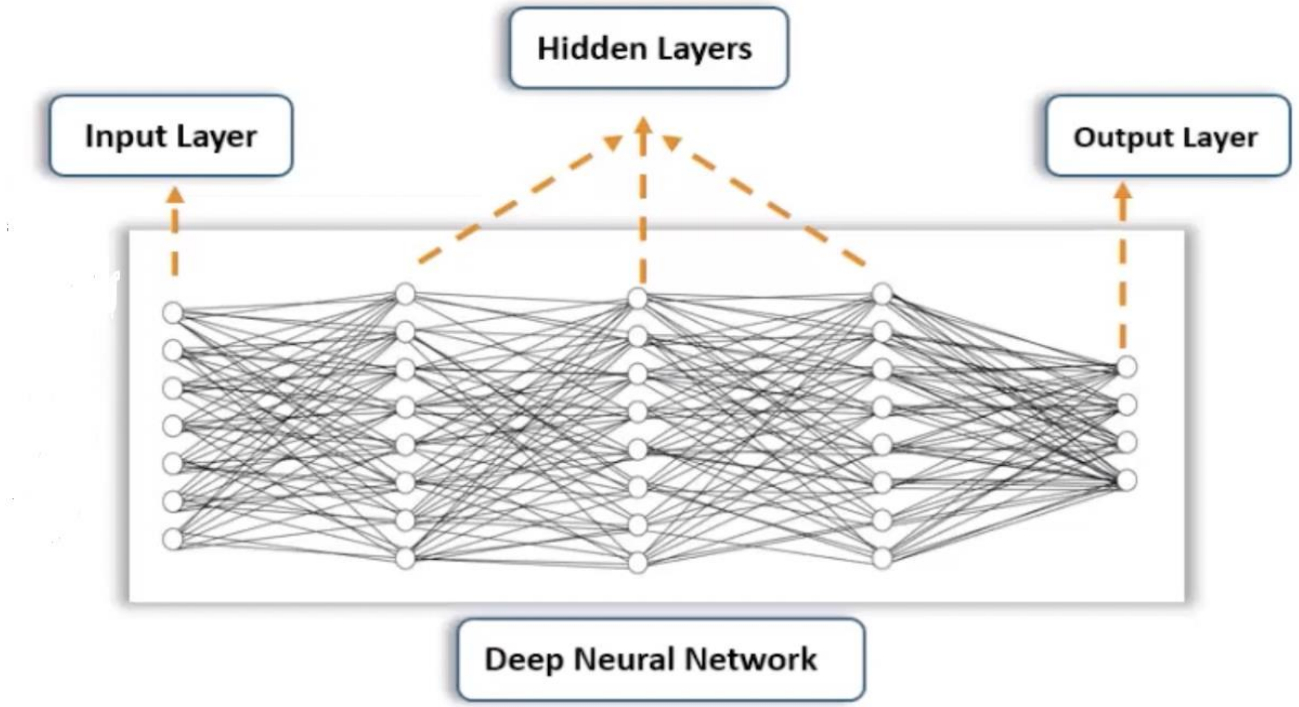
Music composition

Deep Learning (Neural Networks)

- The idea behind deep learning algorithm, you get input from observation and you put your input into one layer
- That layer creates an output which in turn becomes the input for the next layer, and so on
- This happens over and over until your final output

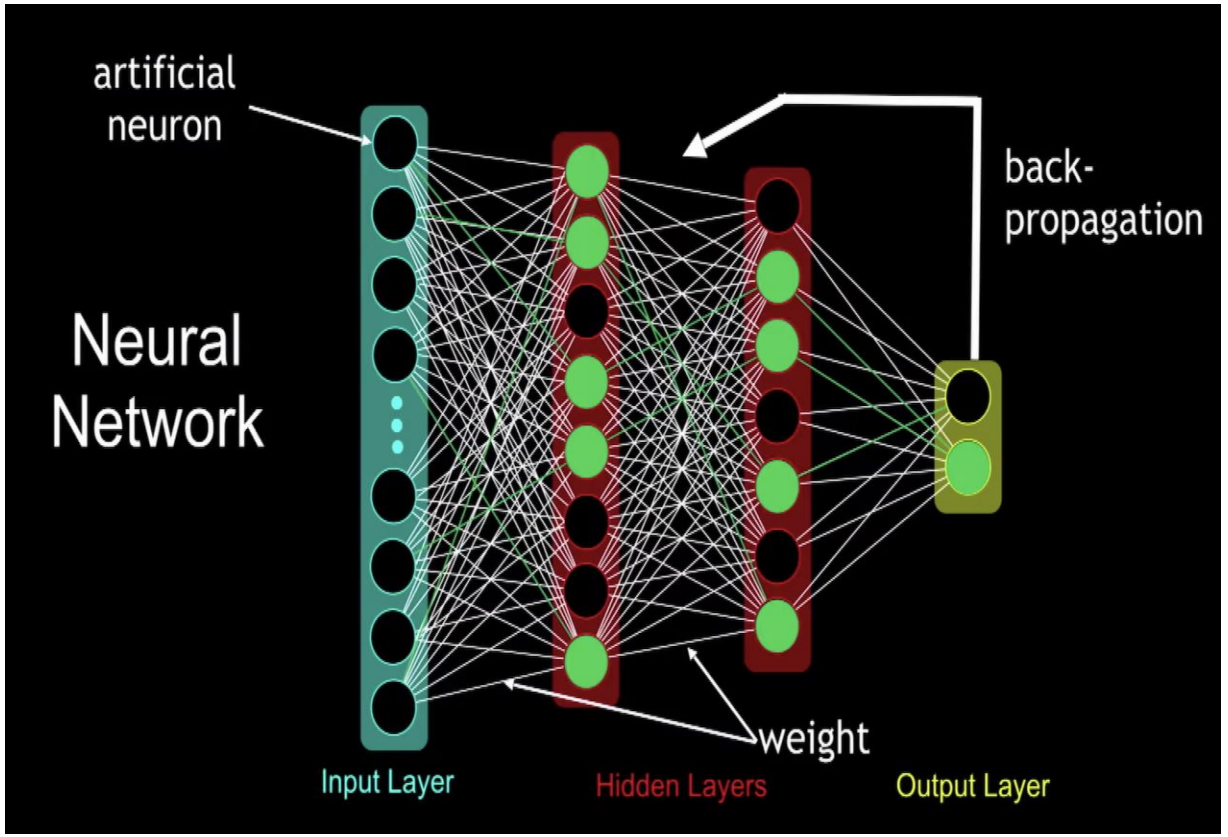
Deep Learning (Neural Networks)

- The network is said to be deeper based on the number of layers it has!



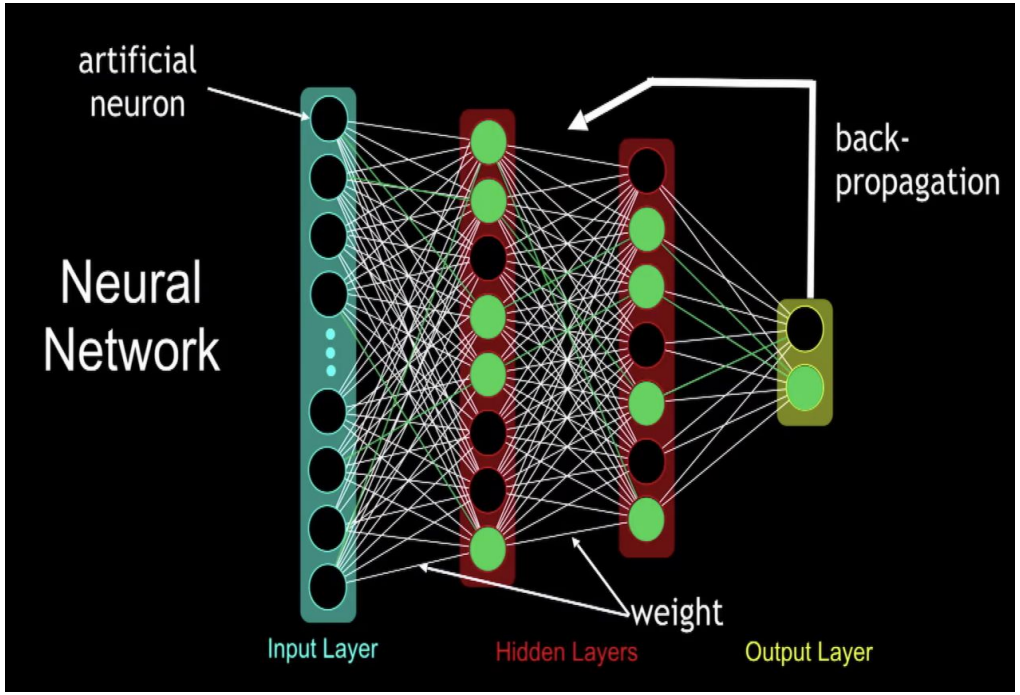
How Neural Networks Work?

Back-propagation and Weight for Deep Neural Networks



How Neural Networks Work?

- Our end goal is to reduce the output error
- Weight is calculated between the layers and adjusted through back-propagation in every iteration

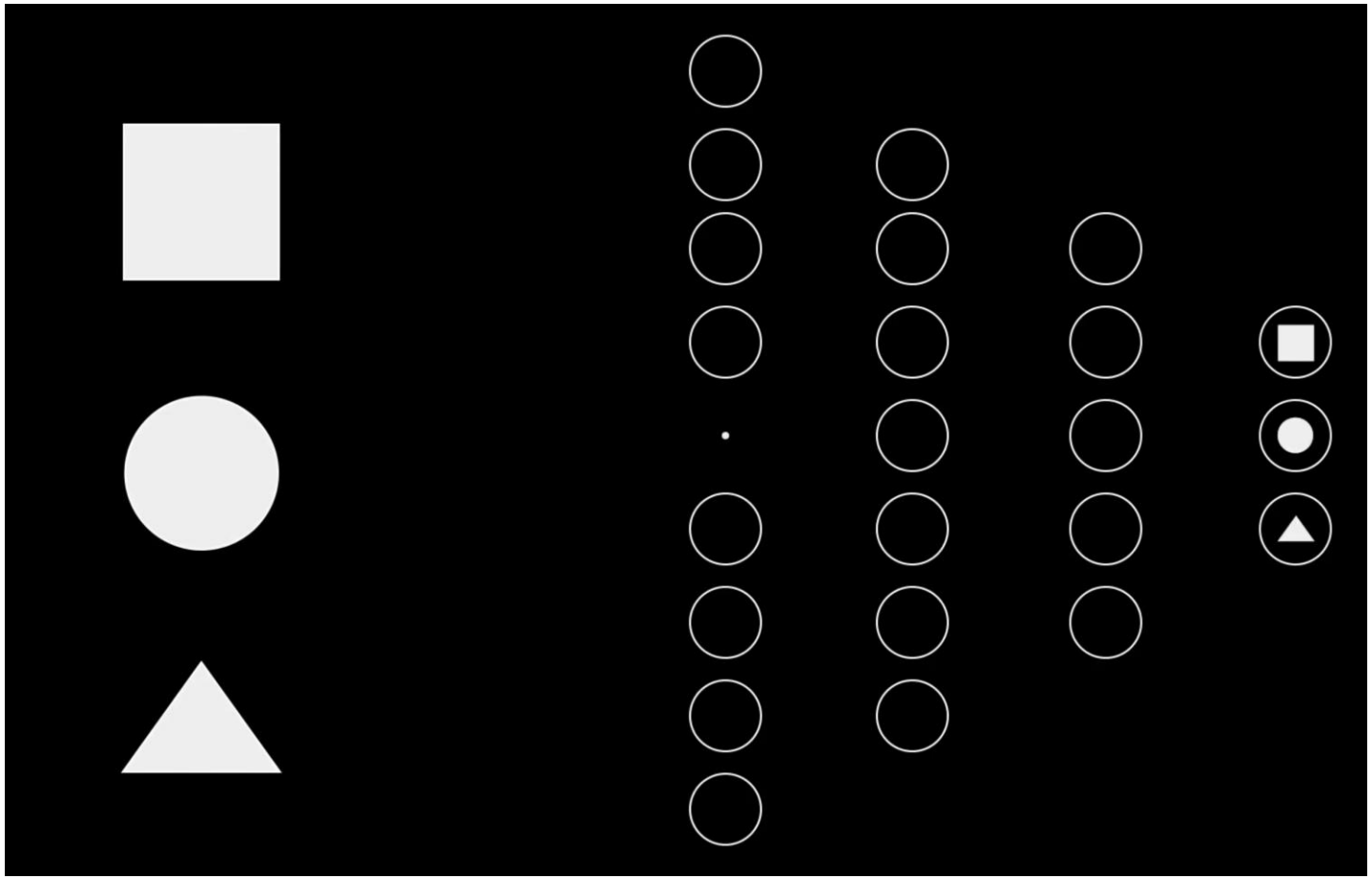


How Neural Networks Work?

- Back-propagation is the essence of neural net training
- It is the practice of fine-tuning the weights of a neural net, based on the error rate obtained in the previous iteration
- Proper tuning of the weights ensures lower error rates

How Neural Networks Work? Example

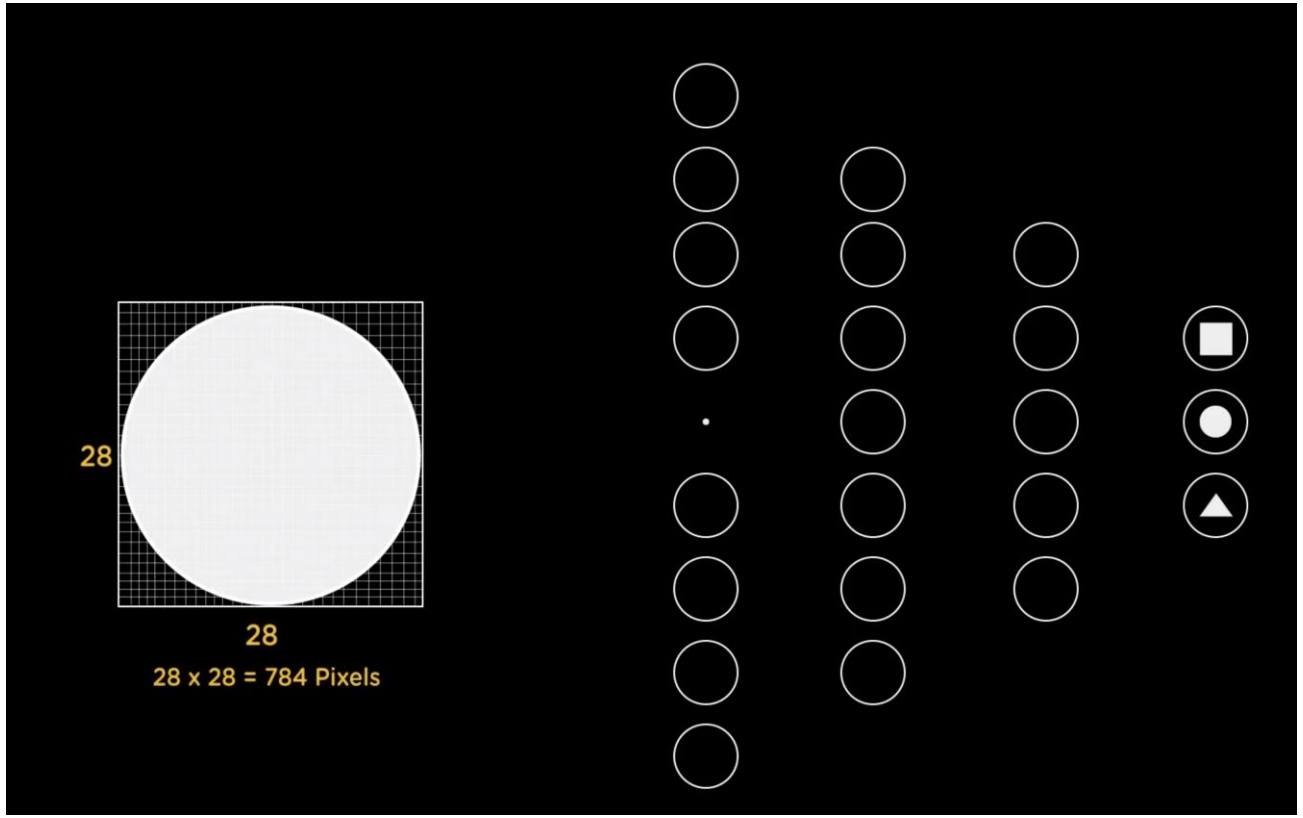
Let's try to identify circle, rectangle and square



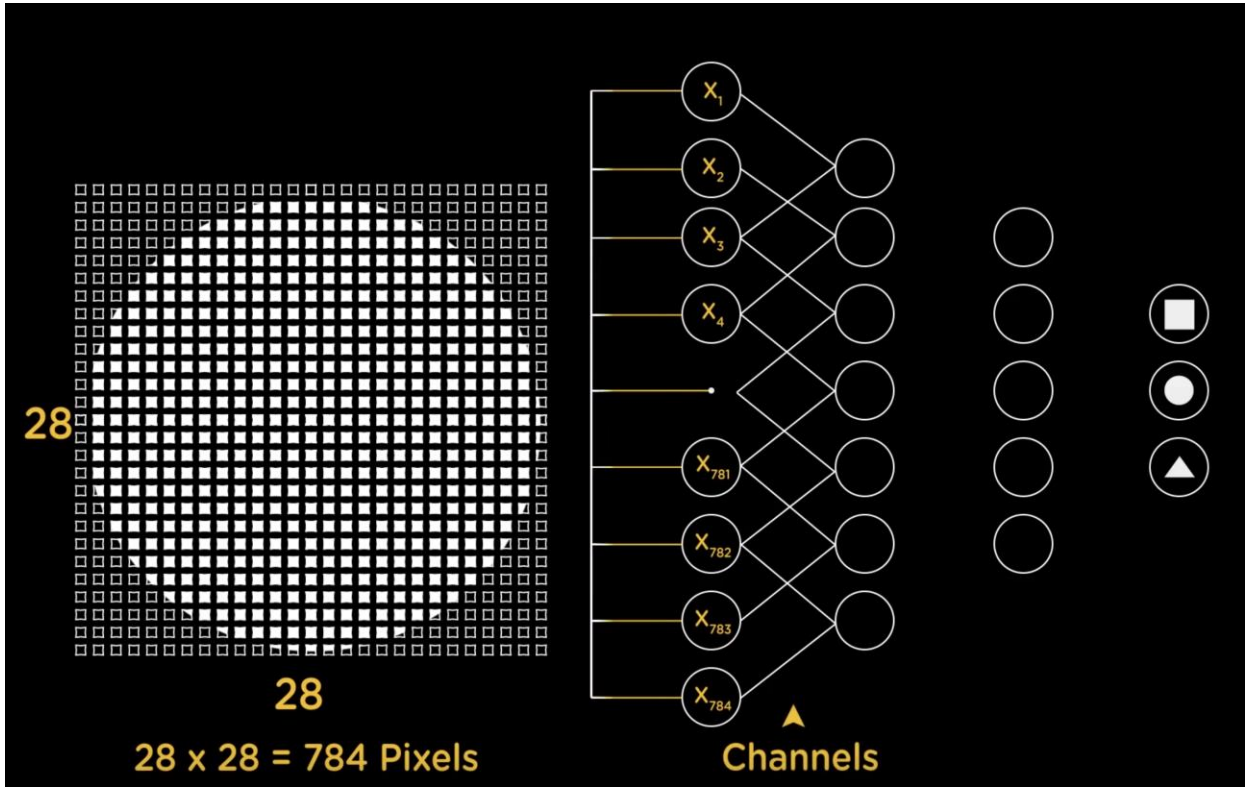
How Neural Networks Work? Example

- Let’s show circle to our neural network and see whether it can correctly identify it

- Circle is 28x28 = 784 pixels and each pixel is placed as input data to the input layer of our neural network



How Neural Networks Work? Example

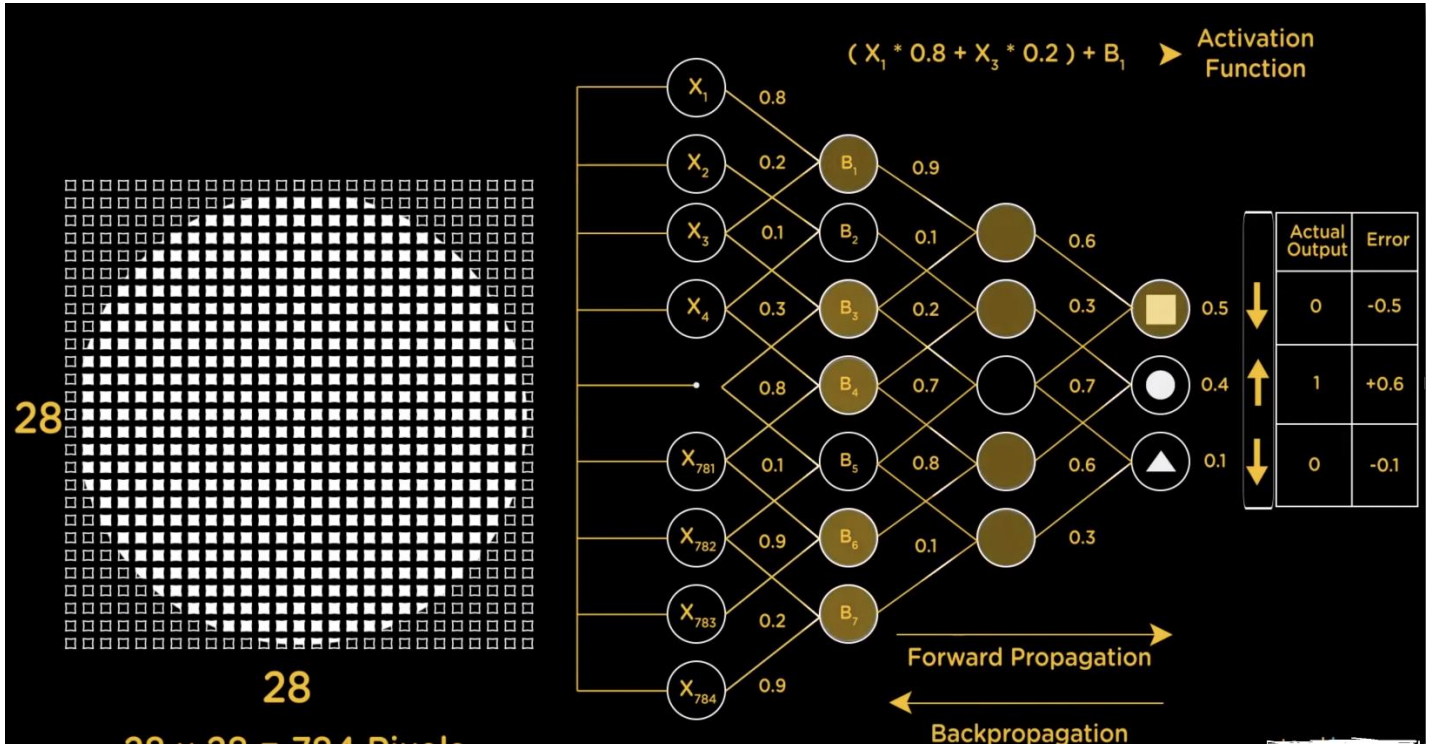


Layers are connected to each other through Channels

How Neural Networks Work? Example

Weight is assigned for each channel initially as random

Both input data and output data is provided to training algorithm during the training process

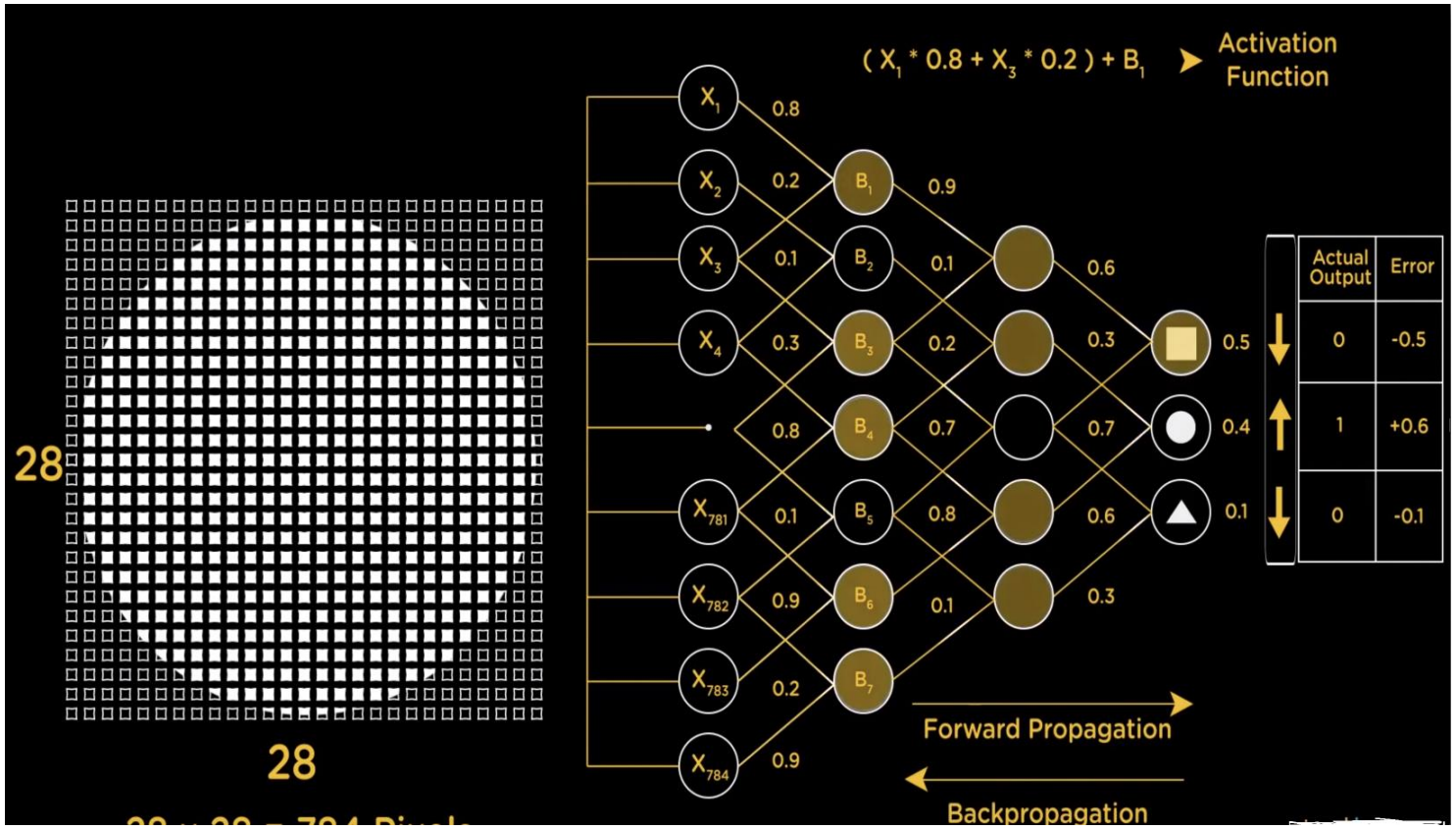


How Neural Networks Work? Example

Activation function is threshold value basically

When the weighted sum exceeds the threshold value, node fires and it will be eligible to send the input to next layer through a forward propagation

When the calculated output – final result is different than actual output, weight is adjusted through back-propagation



Segment Routing

Segment Routing Basics

- Segment Routing (SR) leverages the source routing paradigm. A node steers a packet through an ordered list of instructions, called ‘ segment ‘
- State is kept in the packet header, not on the router, with Segment Routing
- Resources such as the CPU and Memory are saved
- If you have 100 Edge Routers in your network and if you enable MPLS Traffic Edge to Edge, you would have 100×99/2 = 4950 LSP states on your Midpoint LSR. This is prevalent in many MPLS TE enabled network

Segment Routing – Why Segment Routing?

- If you enable Segment Routing and if you evaluate the same midpoint case (since you assign a Prefix/Node SID for every Edge router), Midpoint LSR would have 110 entries instead of 4500 entries

- As for the scalability, everything is perfect. However, there is a caveat

- Segment list can easily get big if you use explicit routing for the purpose of OAM. If you do that, you may end up with 7-8 segments. In that case, it is pertinent that you check the hardware support

Segment Routing – Why Segment Routing?

- Cisco claims that they have performed the tests on a number of service provider networks and that their findings show that two or three segments would be enough for the most explicit path scenarios

- You can use Segment Routing to provide MPLS VPN service without using LDP for the transport label distribution

Segment Routing MPLS – How does it work?

- There are mainly two types of Segment in Segment Routing with MPLS Dataplane : Prefix and Adjacency Segment
- Prefix Segment is used for the shortest-path to the IGP prefix and it is Equal Cost Multi Path (ECMP) aware
- Prefix Segment is distributed by ISIS or OSPF

Segment Routing MPLS – How does it work?

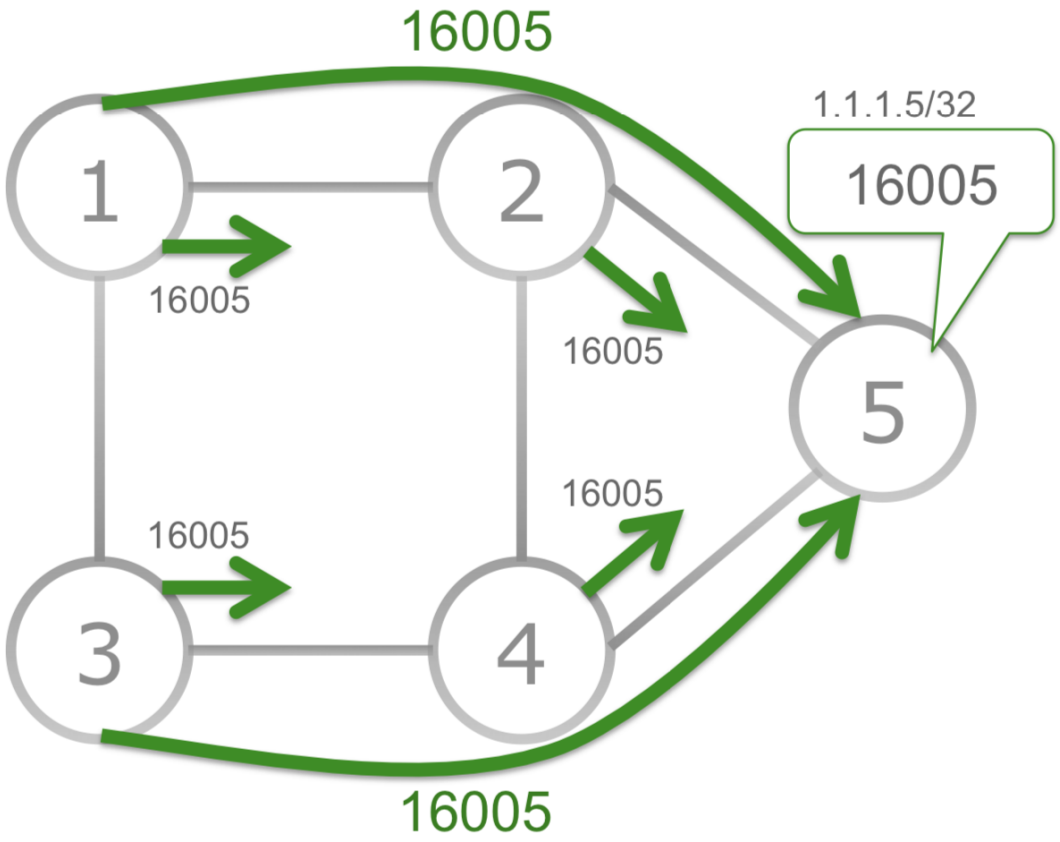
- Adjacency segment is forward on the IGP adjacency
- Adjacency segment is local segment , not domain wide
- Adjacency segment is distributed by ISIS or OSPF as well

Segment Routing MPLS – How does it work?

ECMP to the Node 5

16005 is unique in the IGP domain

16005 is advertised in IGP

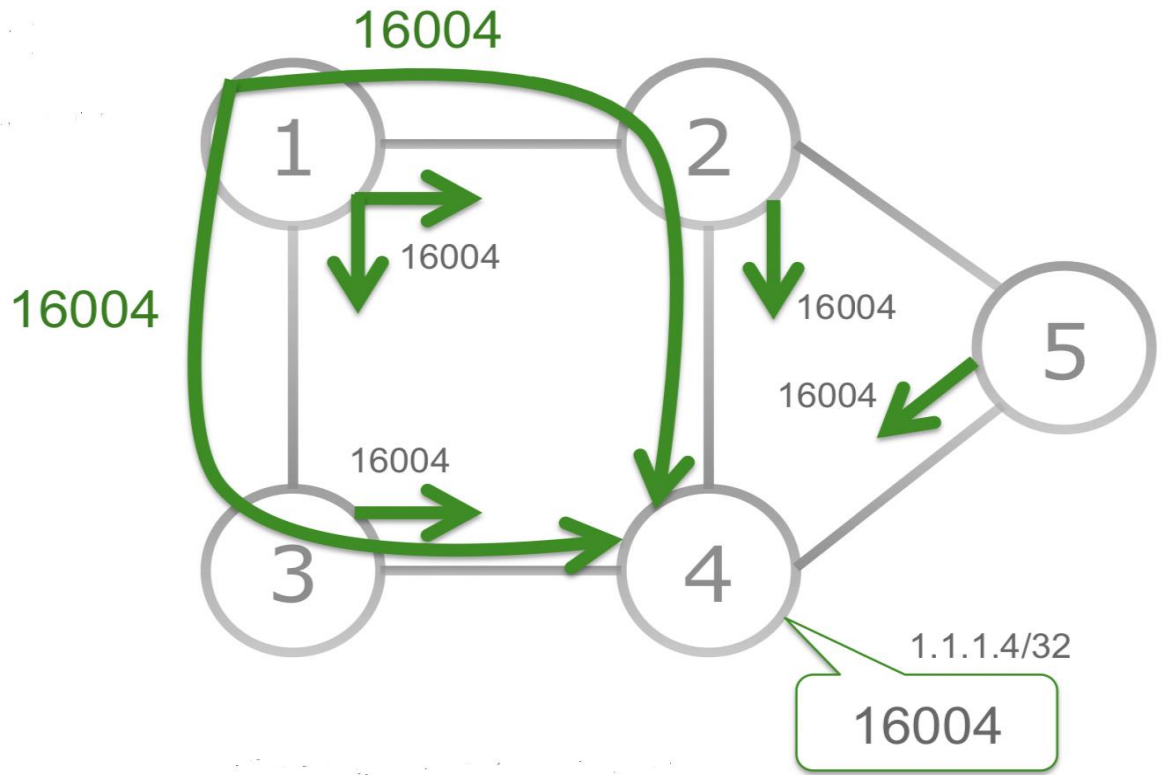


Segment Routing MPLS – How does it work?

ECMP to the Node 4

16004 is unique in the IGP domain

16004 is advertised in IGP

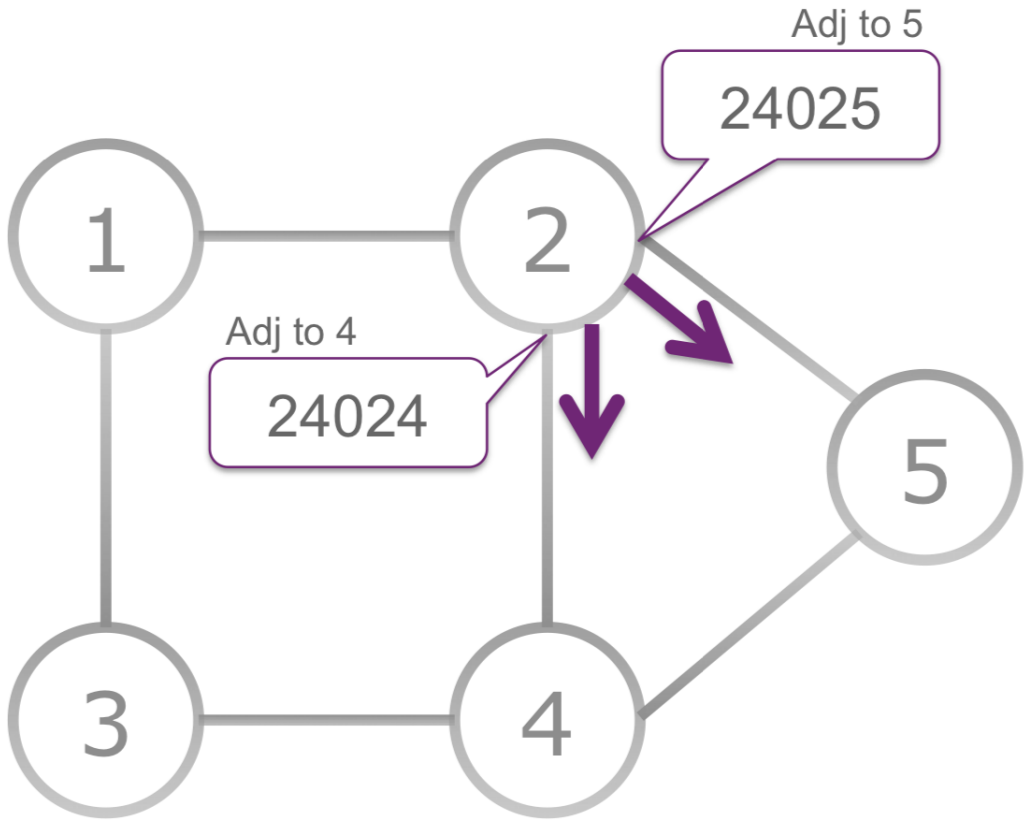


Segment Routing MPLS – How does it work?

24024 and 24025 are Adjacency Segments of Node 2

24024 is Local Adjacency Segment towards Node 4

24025 is Local Adjacency Segment towards Node 5



Segment Routing MPLS – How does it work?

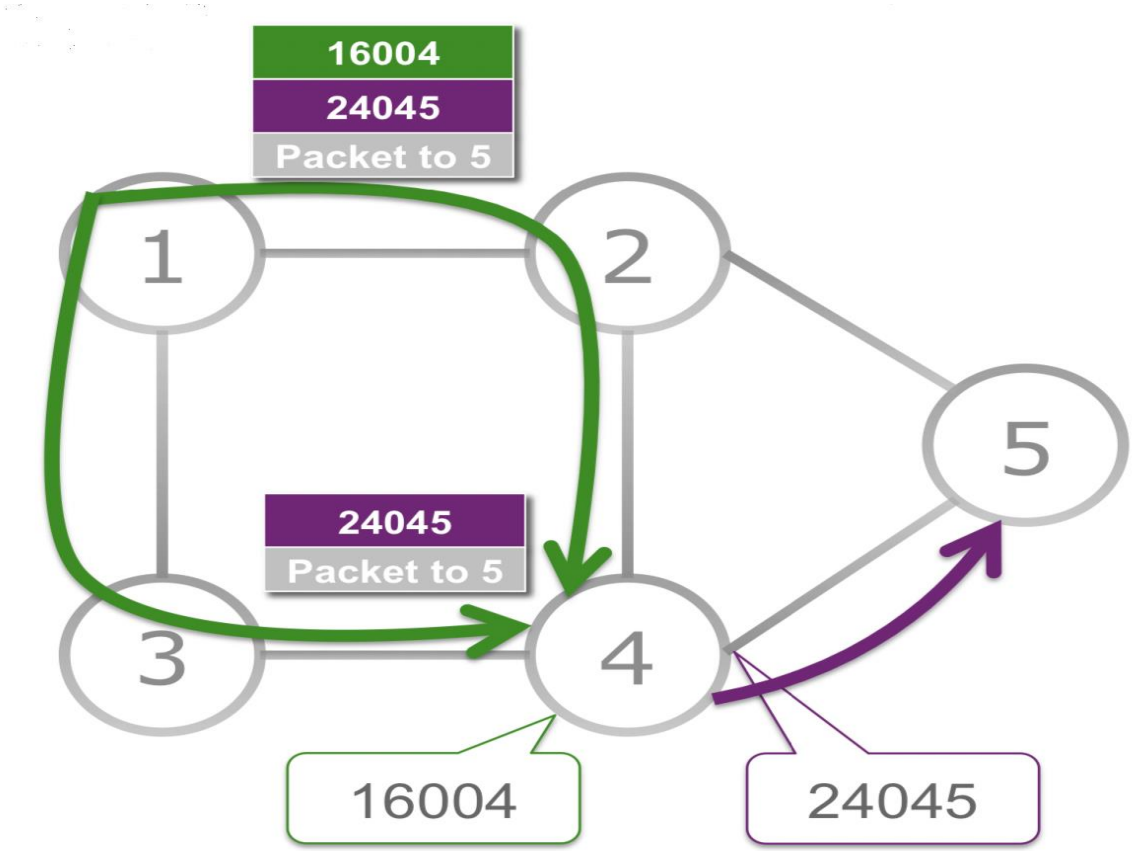
Two Segments encoded

24045 and 16004

16004 is Prefix Segment of Node 4

16004 is used to reach Node 4 in an ECMP manner

24045 is Adjacency Segment of Node 4 towards Node 5 Link



Segment Routing – Traffic Engineering Capability

- Segment Routing provides Traffic Engineering without having soft state RSVP-TE protocol on your network. Soft state protocols require a lot of processing power

- Although Segment Routing does not have permission control, you can use routers to specify, for instance, 50Mbs LSP path for traffic A and 30 Mbps for traffic B using centralized controller, a process that allows you to use traffic engineering

Segment Routing – Use Cases/Applications

- Segment Routing provides Fast Reroute without RSVP-TE, and you do not need to have thousands of forwarding state in the network, as it uses IP FRR technology, specifically Topology Independent LFA

- Segment Routing has many use cases
  1. MPLS VPN
  2. Traffic Engineering
  3. Fast Reroute
  4. Dual Plane topologies are some use cases

Segment Routing – Use Cases/Applications – SR-TE

- With Traffic Engineering, you can have ECMP capability, a task that is very difficult to achieve with RSVP Based Traffic Engineering. This is because you need to create two tunnels

Segment Routing – Use Cases/Applications - EPE

- There are other use cases such as Egress Peer engineering
- Today, this can be achieved by the complex BGP policy or LISP
- However, with Segment Routing, BGP Egress peer engineering is much easier

Segment Routing Current Status

- Major vendors – including Alcatel, Ericson, and Juniper – support segment Routing
- If you have devices not supported by Segment Routing but by LDP, you can use Segment Routing to interwork the LDP enabled devices
- Also, the Segment Routing Mapping Server provides interworking functionality

Segment Routing (SRv6)
SR IPv6 Dataplane

Segment Routing (SRv6) – SR IPv6 Dataplane

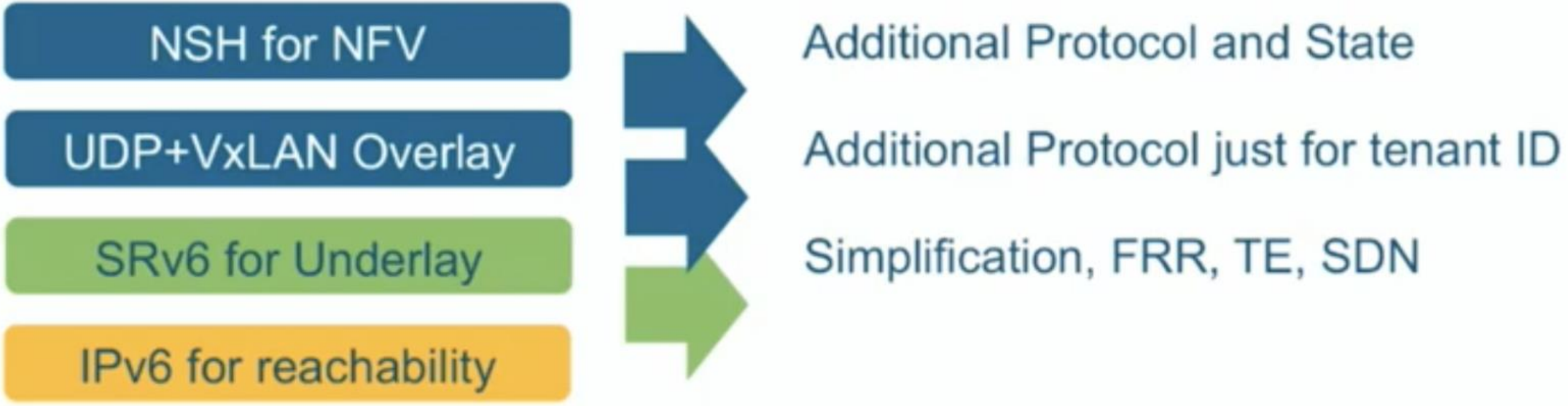
- Segment Routing works based on Source Routing
- Two dataplane is defined for Segment Routing : MPLS andIPv6
- MPLS has been deployed in many networks
- Segment routing is applied to an IPv6 data plane by encoding IPv6 segments into new routing extension header (SRH)

Segment Routing (SRv6) – SR IPv6 Dataplane

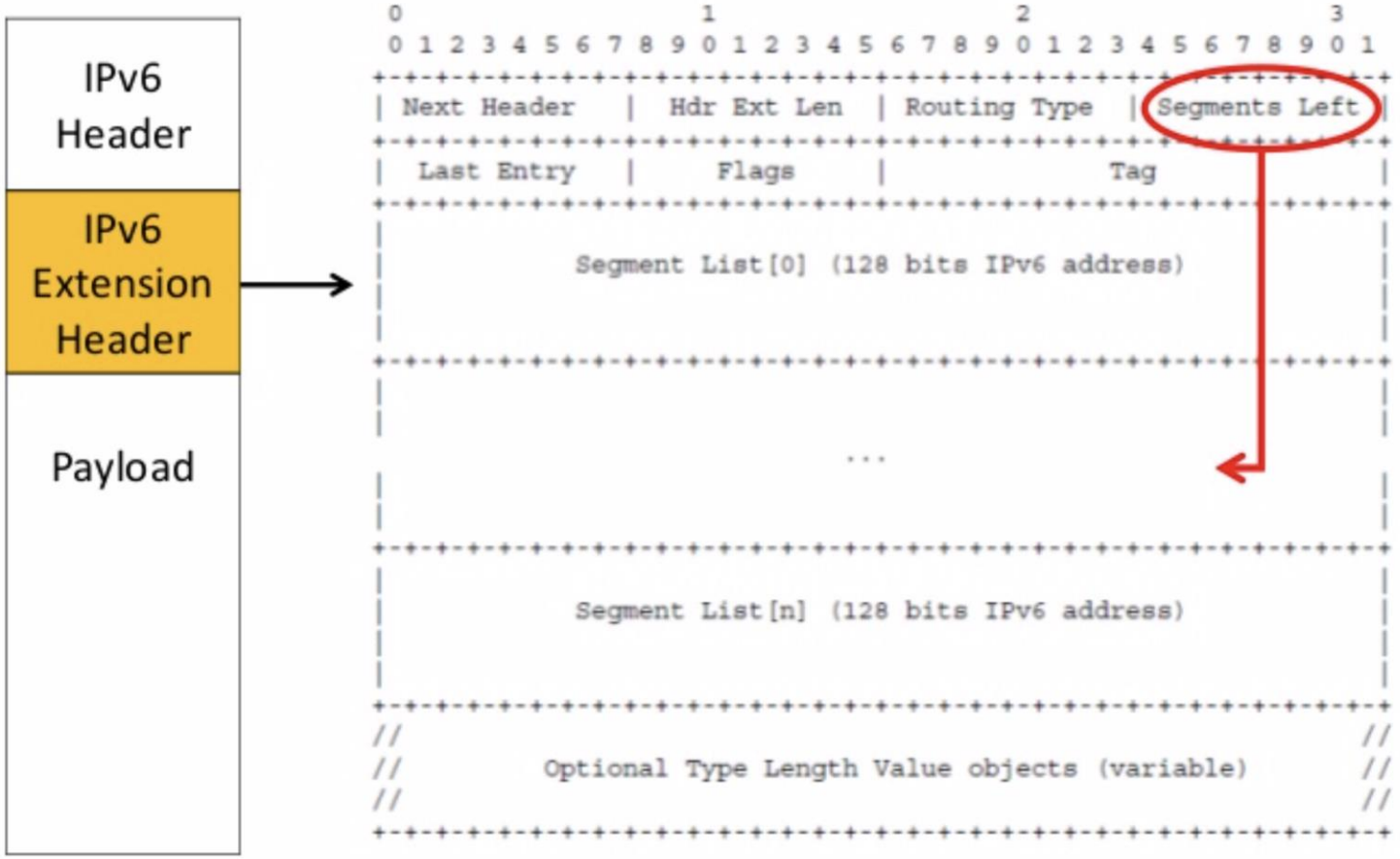
- SR brings Scalability , removes some protocol requirements such as LDP, RSVP , NSH
- Can provide 50msec FRR coverage in any topology with TI-LFA
- Can be used as unified control plane for DC, WAN and Metro Networks
- SR MPLS in DC can be used but generally hosts don’t deploy MPLS, thus SRv6 is seen as better candidate to be deployed towards up to the Host as IPv6 in the host is supported by every vendor

Segment Routing (SRv6) – SR IPv6 Dataplane

- SR MPLS is used for Transport purpose , SRv6 can be used not only for Transport, but also for Service signaling, so much more protocol can be eliminated in the network



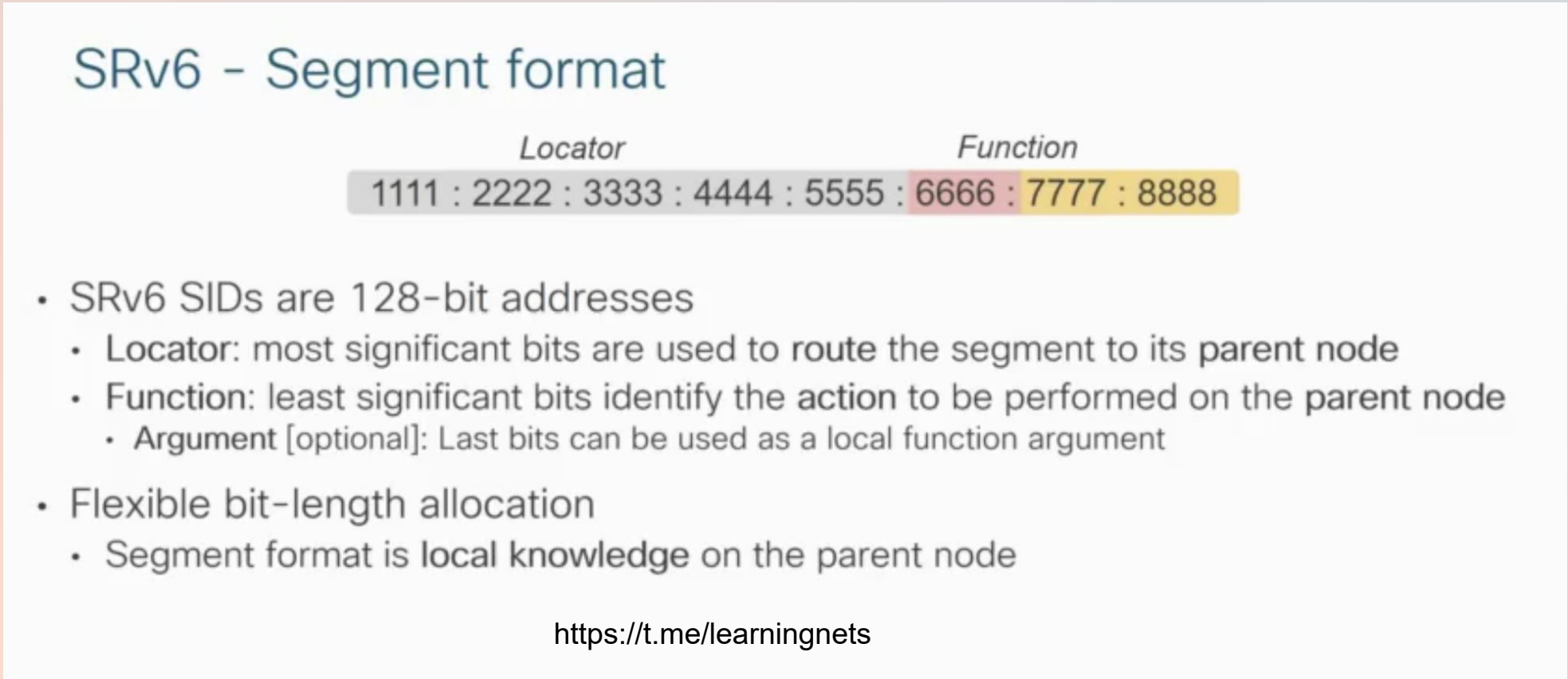
Segment Routing Header in IPv6



- Routing Type
  - 4 (Segment Routing)
- Segments Left
  - Index to the next segment in the Segment List
  - Decremented on Endpoint node
- Last Entry
  - Index to the first segment in the Segment List
- Segment List
  - Encoded starting from the last segment of the path (Segment List [0] contains the last segment)

Segment Routing (SRv6) – SR IPv6 Dataplane

- When SRv6 is deployed, only the nodes that have to process the packet header must have SRv6 dataplane support, all other nodes in the network are just plain IPv6 nodes



Segment Routing (SRv6) – SR IPv6 Dataplane

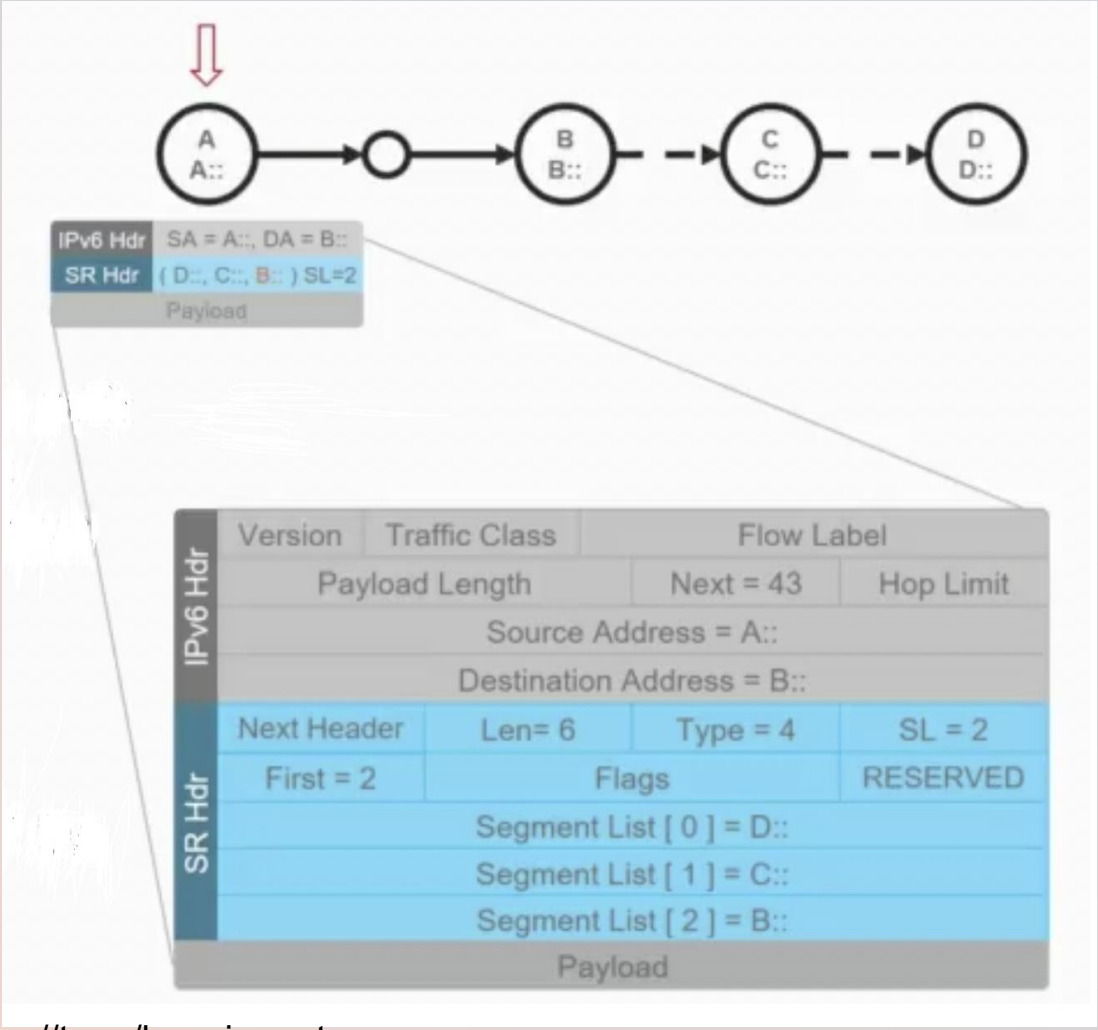
- 128 bit segment ID is broken into three fields: Locator, Function and Arguments
- These represent the forwarding information (Locator) and any actions to be performed at that destination (Function), plus any information required by the individual SID (Arguments)
- The Argument field of the SID could carry the QoS Flow Identifier (QFI), for example

Segment Routing (SRv6) – SR IPv6 Dataplane

- Locator part is routable in an IPv6 network
- Locator information is distributed by IGP and all other nodes install this information to their IPv6 routing table. Even it is not a real address, the other nodes will be able to route packets to the this address (aka SRv6 SID)
- Segment Left (which is decremented at every SRv6 hop) is copied to the IPv6 Destination field, allowing standard routing practices to be applied to SRv6 packets when traversing non-SRv6 capable network elements

Segment Routing (SRv6) – Non-SRv6 Capable Nodes

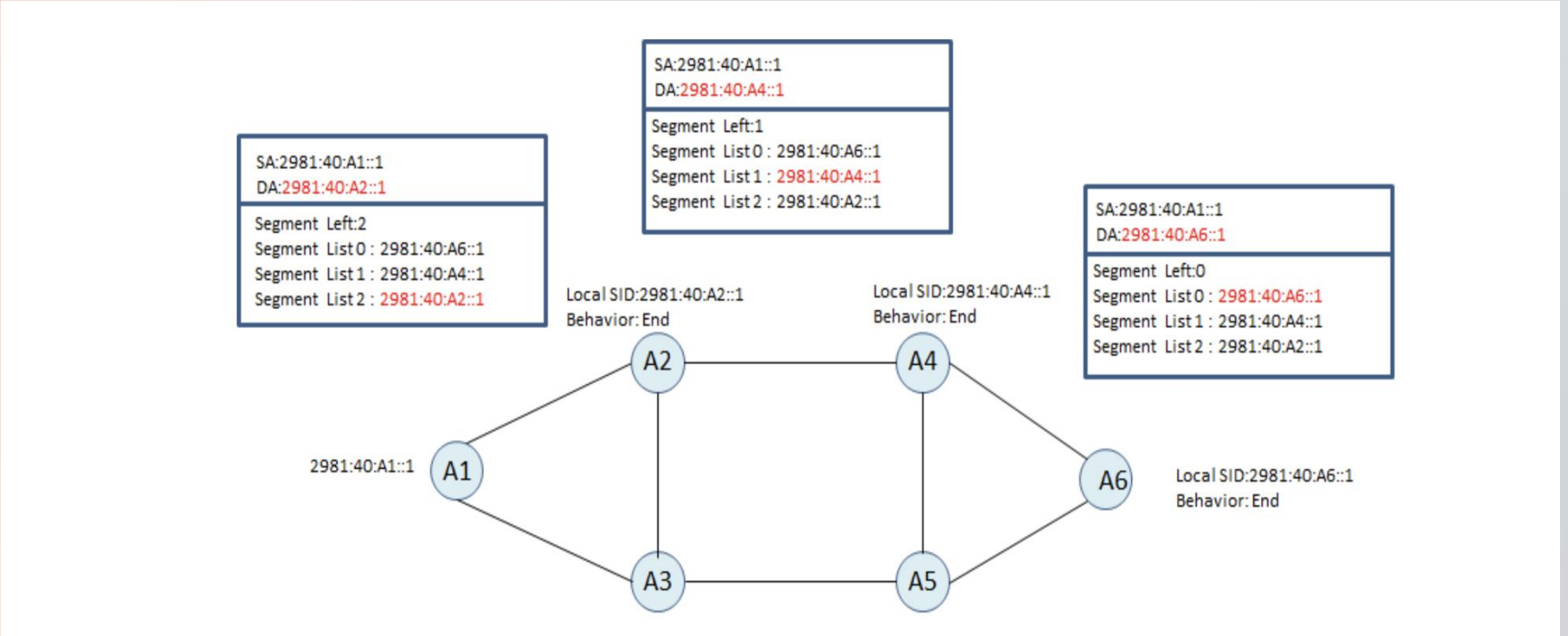
- Non-SRv6 capable nodes just perform IPv6 routing, they don't have to understand or take any action on SRH



SRv6 Basic Functions - END Function

- This is corresponding to a Node SID in SR-MPLS
- When the node receive the packet with End Function (Function 0) it decrements the Segments Left field, update the Destination Address field in the IPv6 header and forward the packet to next node along the shortest path route (Node SID in SR-MPLS uses shortest path tree as well)
- If Segment Left is 0, it means final node is reached, thus IPv6 and SRH headers are removed and payload is processed

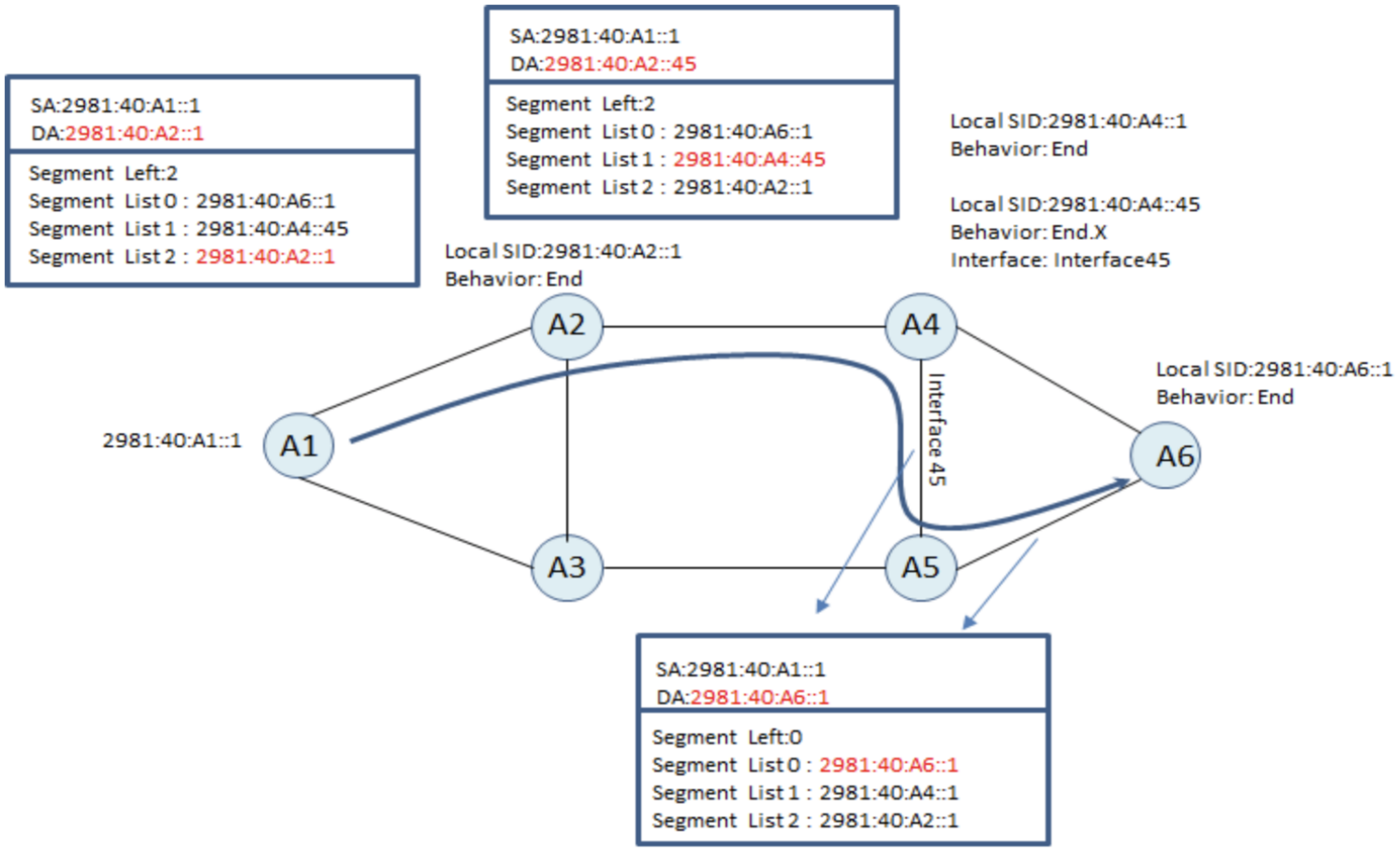
SRv6 Basic Functions - END Function



SRv6 Basic Functions - END.X Function

- End.X function is defined locally and it points an outgoing interface. Official definition for End.X function is Endpoint with cross-connect to layer-3 adjacencies
- It is similar to Adjacency-SID in SR-MPLS
- When traffic needs to be pushed to a certain interface, END.X function is used , Example use case for this function is TI-LFA

SRv6 Basic Functions - END.X Function



SRv6 Functions List

Functoins Defined in Net Programming

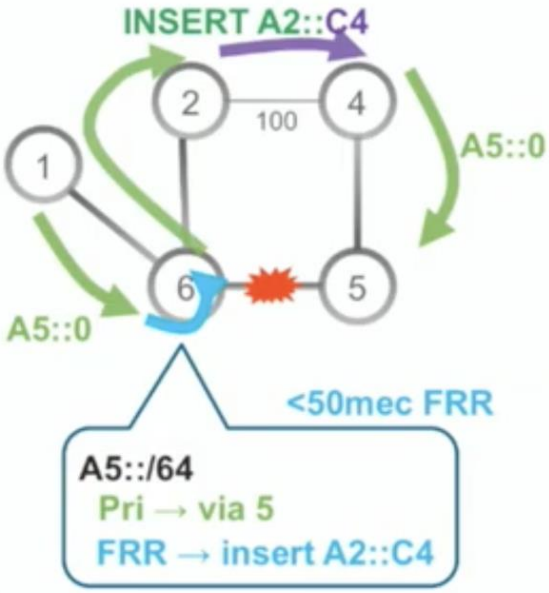
- **End** Endpoint function The SRv6 instantiation of a prefix SID
- **End.X** Endpoint function with Layer-3 cross-connect The SRv6 instantiation of a Adj SID
- **End.T** Endpoint function with specific IPv6 table lookup
- **End.DX2** Endpoint with decapsulation and Layer-2 cross-connect L2VPN use-case
- **End.DX2V** Endpoint with decapsulation and VLAN L2 table lookup EVPN Flexible cross-connect use-cases
- **End.DT2U** Endpoint with decaps and unicast MAC L2 table lookup EVPN Bridging unicast use-cases
- **End.DT2M** Endpoint with decapsulation and L2 table flooding EVPN Bridging BUM use-cases with ESI filtering
- **End.DX6** Endpoint with decapsulation and IPv6 cross-connect IPv6 L3VPN use (equivalent of a per-CE VPN label)
- **End.DX4** Endpoint with decapsulation and IPv4 cross-connect IPv4 L3VPN use (equivalent of a per-CE VPN label)
- **End.DT6** Endpoint with decapsulation and IPv6 table lookup IPv6 L3VPN use (equivalent of a per-VRF VPN label)
- **End.DT4** Endpoint with decapsulation and IPv4 table lookup IPv4 L3VPN use (equivalent of a per-VRF VPN label)
- **End.DT46** Endpoint with decapsulation and IP table lookup IP L3VPN use (equivalent of a per-VRF VPN label)
- **End.B6** Endpoint bound to an SRv6 policy SRv6 instantiation of a Binding SID
- **End.B6.Encaps** Endpoint bound to an SRv6 encapsulation Policy SRv6 instantiation of a Binding SID
- **End.BM** Endpoint bound to an SR-MPLS Policy SRv6/SR-MPLS instantiation of a Binding SID
- **End.S** Endpoint in search of a target in table T

- **T.Insert** Transit behavior with insertion of an SRv6 policy
- **T.Insert.Red** Transit behavior with reduced insert of an SRv6 policy
- **T.Encaps** Transit behavior with encapsulation in an SRv6 policy
- **T.Encaps.Red** Transit behavior with reduced encaps in an SRv6 policy
- **T.Encaps.L2** T.Encaps behavior of the received L2 frame
- **T.Encaps.L2.Red** Transit with reduce encaps of received L2 frame

TI-LFA with SRv6 Use Case

TILFA

- 50msec Protection upon
  local link, node or SRLG failure
- Simple to operate and understand
  - automatically computed by the router's IGP process
  - 100% coverage across any topology
  - predictable (backup = post convergence)
- Optimum backup path
  - leverages the post-convergence path, planned to carry the traffic
  - avoid any intermediate flap via alternate path
- Incremental deployment
- Distributed and Automated Intelligence



VPNv4 example with SRv6 Dataplane