Design Guide CX - Wireless for Large Public Networks

Contents

- **Introduction**
  - CX Design Guide
  - Scope and Definitions
  - Large Public Networks
  - External References
  - Disclaimer
- **Designing the Network**
  - RF considerations
    - Venue types
    - Coverage Strategies
    - Aesthetics
    - Rogue networks
    - Single 5GHz vs. Dual 5GHz
    - Antennas
    - High Density and 6GHz
    - Radio Resource Management
  - RF Configuration
    - Channels
    - Data rates
    - Transmit Power
    - Power Balance
    - RxSOP
  - Scaling the network
    - Number of APs
    - WLC Platform
    - WLC High-Availability
    - External Systems
    - DNS/DHCP
- **Operating the Network**
  - The Right Configuration
  - SSIDs
    - How many SSIDs?
    - WPA2/3 Personal
    - WPA2/3 Enterprise
    - Guest SSIDs
    - Conclusion on the number of SSIDs
    - The Legacy SSID versus Main SSID concepts
    - SSID features
  - Site Tag
  - Policy Profile
  - AP Join Profile

- Monitoring the Network
  - The harder you monitor, the more you can cause your own problems
  - Issues Specific to Large Networks
  - Day 2 Monitoring: Keeping an Eye on User Satisfaction
- Configuring for Scalability
  - SVIs and Interfaces on the 9800
  - Aggregated Probe Response
  - IPv6
  - mDNS
- Hardening the Network
  - Security
  - Rogue Access Points
  - WiPS
  - Rate limiting clients
  - Restricting Client Access
  - Protecting from Traffic Storms

Conclusion

Introduction

This document describes design and configuration guidelines for large public Wi-Fi Networks.

CX Design Guide



CX Design Guides are written by specialists from Cisco Technical Assistance Center (TAC) and Cisco Professional Services (PS) and peer-reviewed by experts within Cisco; the guides are based on Cisco leading practices as well as knowledge and experience gained from countless customer implementations over many years. Networks designed and configured in line with the recommendations in this document help avoid common pitfalls and improve network operation.

Scope and Definitions

This document provides design and configuration guidelines for large public wireless networks.

Definition: Large public networks - wireless deployments, often at high-density, that provide network connectivity for thousands of unknown and/or unmanaged client devices.

This document often assumes the target network is providing services to large and/or temporary events. It

also fits static permanent networks for venues that receive many guests. For example, a shopping mall or airport have similarities with the Wi-Fi network of a stadium or concert venue - in the sense that there is no control over end users, and they exist in the network typically just for a couple of hours, or for the day at most.

Wireless coverage for large events or venues has its own set of requirements, which tends to be different from enterprise, manufacturing, or even large education networks. Large public networks can have thousands of people, concentrated in just one or a few buildings. They can have very frequent client roaming, constantly or during peaks, plus the network must be as compatible as possible with *anything* in terms of wireless client devices, with no control over client device configuration or security.

This guide presents general RF concepts for high-density as well as implementation details. Many of the radio concepts in this guide apply to all high-density networks, regardless of hardware platform. However, implementation details and configurations are focused on Catalyst Wireless using the Catalyst 9800 Wireless Controller, as this is the most common solution deployed for large public networks today.

This document uses the terms Wireless Controller and Wireless LAN Controller (WLC) interchangeably.

Large Public Networks

Large public and event networks are unique in many aspects, this document explores and provides guidance on these key areas.

- Large public networks are intense; there are thousands of devices in a reduced Radio Frequency (RF) space and significant roaming as people walk around, some events and venues can be more static with bandwidth peaks at very specific times. The infrastructure needs to handle all these state changes as gracefully as possible for clients that enter and move around the area.
- Ease of onboarding is a priority. An associated client is a happy client. This means you want to make the client association to the network as fast as possible. A client that is not connected to Wi-Fi scans for available access points which generates unwanted RF energy, which translates to additional congestion and lost capacity over the air.
- The RF deployment needs to be designed as carefully as possible. A proper RF design using directional antennas is a must if very high density is required, or if the venue has large open spaces and/or high ceilings.
- Another key design drive is compatibility. Some features are standard in the 802.11 specification while other features are proprietary, neither poses any problem to clients. However, reality is different and there are many poorly programmed client drivers that misbehave when faced with complicated beacons or features/settings they do not understand.
- Troubleshooting is challenging due to scale and time restrictions. If something does not work with a specific client, you are not able to work with that end-user to understand the problem. Users can be difficult to find but also can be non-cooperative due to the transient nature of their visit in the venue.
- Security is an important factor. There is less control due to the very large amount of guest visitors and a much bigger attack surface.

External References

Document Name	Source	Location
Cisco Catalyst 9800 Series Configuration Best Practices	Cisco	Link
Troubleshoot Wireless LAN Controller CPU	Cisco	Link
Validate Wi-Fi Throughput: Testing and Monitoring Guide	Cisco	Link
Cisco Catalyst CW9166D1 Access Point Deployment Guide	Cisco	Link
Cisco 9179F deployment guide	Cisco	link
Catalyst 9104 Stadium Antenna (C-ANT9104) Deployment Guide	Cisco	Link
Monitor Catalyst 9800 KPIs (Key Performance Indicators)	Cisco	Link
Troubleshoot Catalyst 9800 Client Connectivity Issues Flow	Cisco	Link
Cisco Catalyst 9800 Series Wireless Controller Software Configuration Guide	Cisco	Link
Wi-Fi 6E: The Next Great Chapter in Wi-Fi White Paper	Cisco	Link

Disclaimer

This document offers recommendations based on certain scenarios, assumptions, and knowledge gained from numerous deployments. However, you the reader are responsible for determining the network design, business, regulatory compliance, security, privacy, and other requirements, including whether to follow the guidance or recommendations provided in this guide.

Designing the Network

RF considerations

Venue types

This guide focuses on large guest networks, typically open to the public, and with limited control over end users and client device types. These types of networks can be deployed in a variety of locations and can be

temporary or permanent. The primary use case is usually providing internet access to visitors, although this is rarely the only use case.

Typical locations:

- Stadiums and arenas
- Conference venues
- Large auditoriums

From an RF point of view each of these location types has its own set of nuances. Most of these examples are usually permanent installations, apart from conference/concert venues, as these can be permanent or set up for a specific show on a temporary basis.

Other locations:

- Cruise ship
- Airport
- Shopping Center / Mall

Airports and cruise ships are also examples of deployments that fit into the category of large public networks; however, these have additional considerations specific to each case and often make use of internal omnidirectional APs.

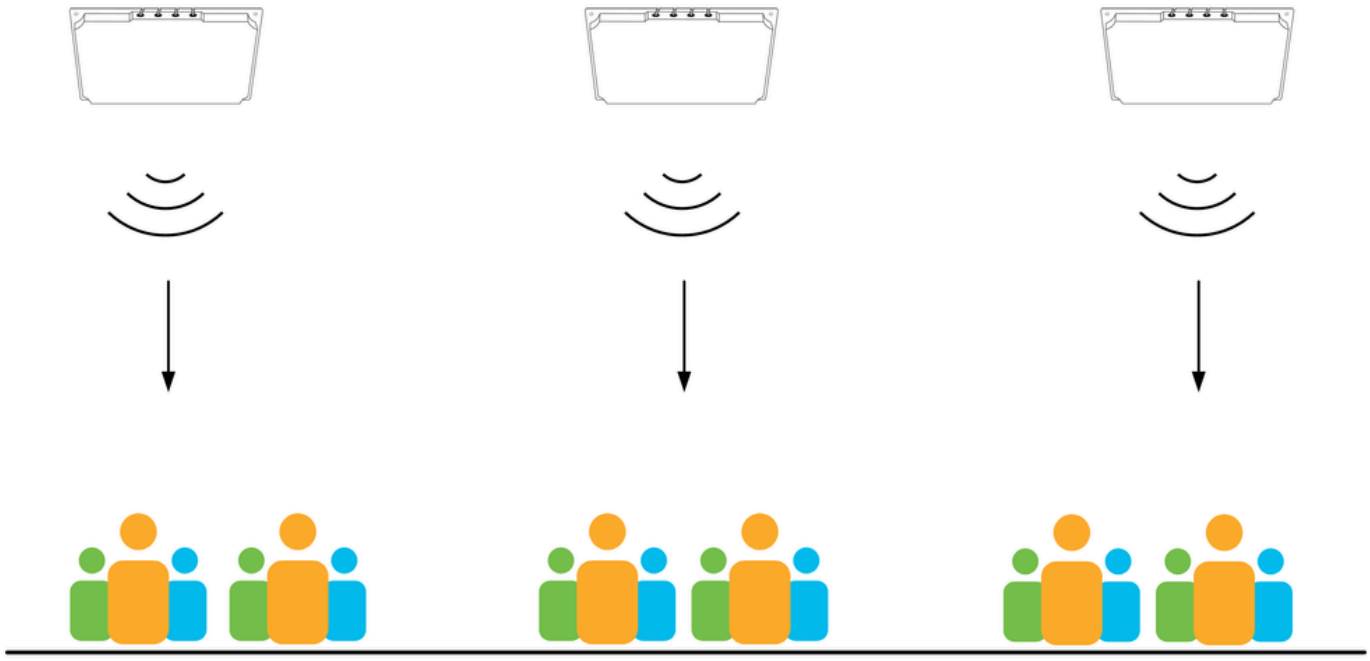
Coverage Strategies

Coverage strategies depend largely on the venue type, antennas used, and available antenna mounting locations.

Overhead

Overhead coverage is always preferred wherever possible.

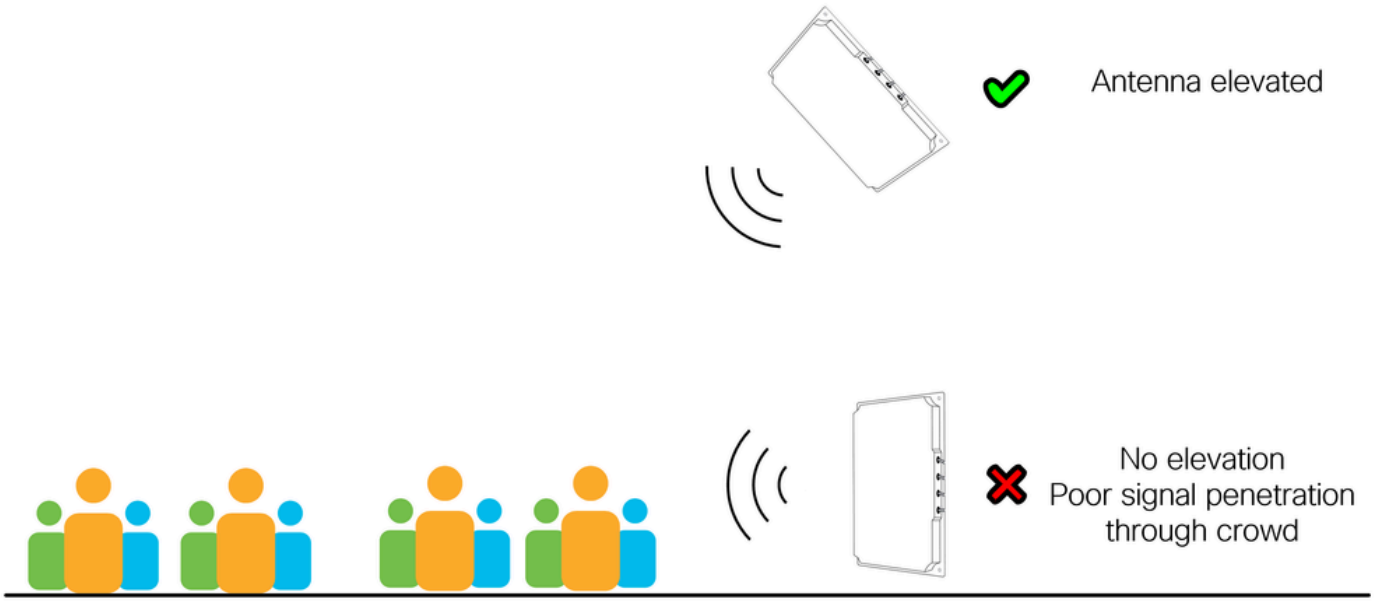
Overhead solutions have the distinct advantage that all client devices typically have direct line of sight to the antenna overhead, even in crowded scenarios. Overhead solutions using directional antennas provide a more controlled and well-defined coverage area making them less complicated from a radio tuning point of view, while providing superior load balancing and client roaming characteristics. See section on Power Balance for additional information.



APs above the clients

Side

Side mounted directional antennas are a popular choice and work well in a variety of scenarios, particularly when overhead mounting is not possible due to height or mounting restrictions. When using side mounting it is important to understand the type of area being covered by the antenna, for example is it an open outdoor area or dense indoor area? If the coverage area is a high-density area with many people, then the antenna *must* be elevated as much as possible as signal propagation through a human crowd is always poor. Remember most mobile devices are used lower down at waist level, not above the user's head! The height of the antenna is less significant when the coverage area is a lower density area.



Antenna elevation is always better

Omnidirectional

The use of omnidirectional antennas (internal or external) must generally be avoided in very high-density scenarios, this is due to the potentially high area of impact for co-channel interference. Omnidirectional antennas must not be used at a height above 6m (does not apply to high-gain outdoor units).

Under-seat

In some arenas or stadiums there can be situations where there are no suitable antenna mounting locations. The last remaining alternative is to provide coverage from below by positioning APs under the seats where users are sitting. This type of solution is more difficult to deploy correctly and is usually more costly requiring significantly more APs and specific installation procedures.

The main challenge with under-seat deployment is the large difference in coverage between a full venue and an empty venue. A human body is very efficient at attenuating radio signal, meaning that when there is a crowd of people surrounding the AP the resulting coverage is significantly smaller as compared to when those people are not there. This human crowd attenuation factor allows for more APs to be deployed which can increase overall capacity. However, when the venue is empty there is no attenuation from the human bodies and significant interference, and this leads to complications when the venue is *partially* full.



Note: Under-seat deployment is a valid but uncommon solution, it must be evaluated on a case-by-case basis. Under-seat deployment is not discussed further in this document.

Aesthetics

In some deployments the question of aesthetics comes into play. These can be areas with specific architectural designs, historical value, or spaces where advertising and/or branding dictates where equipment can (or cannot) be mounted. Specific solutions can potentially be required to work around any placement limitations. Some of these workarounds include hiding the AP/antenna, covering the AP/antenna, mounting the equipment in an enclosure, or simply using a different location. Painting the antenna however voids the warranty. Cisco generally does not sell enclosures for antennas, but many are easily available through various providers who also sell various covers and ways to make access points match various aesthetics.

All such workarounds have an impact on the performance of the network. Wireless architects always start by proposing optimal mounting positions for best radio coverage, and these initial positions usually provide the best performance. Any changes to these positions often result in antennas being moved away from their optimal locations.

Locations where antennas are mounted are often elevated, these can be ceilings, catwalks, roof structures, beams, walkways, and any location that provides some elevation over the intended coverage area. These locations are usually shared with other installations, such as: audio equipment, air-conditioning, lighting, cameras and various detectors / sensors. As an example, audio and lighting equipment must be mounted in very specific locations - but why is this? Simply, it is because audio and lighting equipment does not work

correctly when it is hidden in a box or behind a wall.

The same applies to wireless antennas, they work best when there is a line of sight to the wireless client device. Prioritizing aesthetics can (and very often does) have a negative effect on wireless performance, diminishing the value of the investment in infrastructure.

Rogue networks

Rogue Wi-Fi networks are wireless networks that share a common RF space but are not managed by the same operator. These can be temporary or permanent and include infrastructure devices (APs) and personal devices (such as mobile phones sharing a Wi-Fi hotspot). Rogue Wi-Fi networks are a source of interference and in some cases also a security risk. The impact of rogues on wireless performance must not be underestimated. Wi-Fi transmissions are limited to a relatively small range of radio spectrum that is shared between all Wi-Fi devices, any misbehaving devices in proximity have the potential to disrupt network performance for many users.

Within the context of large public networks, these are usually carefully designed and tuned using specialized antennas. A good RF design covers only the areas that are required, often using directional antennas with tuned send and receive characteristics for maximum efficiency.

On the other end of the spectrum are consumer grade devices, or devices supplied by internet service providers. These either have limited options for fine RF adjustment, or are configured for maximum range and perceived performance, often with high power, low data rates, and wide channels. The introduction of such devices into a large event network has the potential to create havoc.

What can be done?

In the case personal hotspots there is very little that can be done, as it would be nearly impossible to monitor tens of thousands of people entering a venue. In the case of infrastructure, or semi-permanent devices, there are some options. Possible remediation starts from simple education, signage for awareness, through signed radio policy documents, ending with active enforcement and spectrum analysis. In all cases, a business decision must be taken on the protection of radio spectrum within the given venue, along with concrete steps to enforce that business decision.

The security aspect of rogue networks comes into play when devices controlled by a 3rd party advertise the same SSID as the managed network. This is equivalent to a honeypot attack and can be used as a method to steal user credentials. It is *always* recommended to create a rogue rule to alert on the detection of infrastructure SSIDs advertised by unmanaged devices. The security section discusses rogues in more detail.

Single 5GHz vs. Dual 5GHz

Dual 5GHz refers to the use of both 5GHz radios on supported APs. There is a key difference between dual 5GHz using external antennas and dual 5GHz using internal antennas (for example, micro/macro cells on omnidirectional APs). In the case of external antennas dual 5GHz is often a useful mechanism, providing additional coverage and capacity while reducing total AP count.

Micro/Macro/Meso

Internal APs have both antennas close together (inside the AP) and are subject to restrictions relating to maximum Tx power when using dual 5GHz. The second radio is limited to low Tx power (this is enforced by the wireless controller) leading to a large imbalance of Tx power between the radios. This can cause the primary (higher power) radio to attract many clients while the secondary (lower power) radio is underutilized. In this case the second radio is adding energy into the environment without providing benefit to clients. If this scenario is observed, it can be better to disable the second radio and simply add another (single 5GHz) AP if additional capacity is required.

Different AP models have different configuration options, the second 5GHz radio can operate at higher power levels in newer *macro/meso* APs like the 9130 and 9136, and some internal Wi-Fi 6E APs like the 9160 series can even operate in *macro/macro*. Always verify the capability of your exact AP model. The second 5GHz slot is also limited in its channel usage, when one slot is operating in one UNII band, the other slot is restricted to a different UNII band, which impacts channel planning and subsequently also available transmit power. Always consider the Tx power difference between dual 5GHz radios, this is true in all cases, including APs with internal antennas.

FRA

Flexible Radio Assignment (FRA) was introduced as a technology to improve 5GHz coverage by switching additional 2.4GHz radios into 5GHz mode, or potentially unused 5GHz radios into monitor mode (for APs that supported it). As this document covers large public networks the assumption is made that the coverage areas as well as radio design are well defined using directional antennas, therefore a deterministic configuration is preferred over a dynamic one. The use of FRA is not recommended for large public networks.

Optionally, FRA could be used when the network is set up to help determine which radios to convert to 5GHz, but once you are happy with the result, it is advised to freeze FRA.

Regulatory

Each regulatory domain defines what channels are available for use and their maximum power levels, there are also restrictions on what channels can be used indoors versus outdoors. Depending on the regulatory domain, it can sometimes not be possible to utilize a dual 5GHz solution effectively. An example of this is the ETSI domain where 30dBm is allowed on UNII-2e channels, but only 23dBm on UNII1/2. In this example, if the design requires the use of 30dBm (usually due to higher distance to the antenna) the use of a single 5GHz radio can be the only feasible solution.

Sometimes extra channels are added in a later software release to certain regulatory domains, but it is possible that those channels have a different maximum transmit power as well : verify that your clients support these channels and only use them on APs where the transmit power will not be a problem.
In case you want to use all the channels available, it can be interesting to create RF profiles where the APs requiring the highest transmit power (such as APs mounted on very high ceiling) only use channels where higher power is available and keep channels of lower power to APs that are closer to the ground in denser areas.

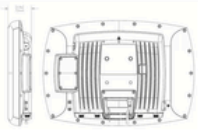
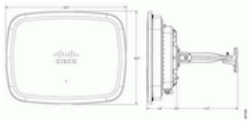
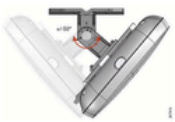
Antennas

Large public networks can use any type of antenna, and typically deploy the most suitable antenna for the job. Mixing antennas within the same coverage area makes the radio design process more challenging and must be avoided if possible. However, large public networks often have large coverage areas with different mounting options even within the same area, making it necessary to mix antennas in some cases.

This table lists the most commonly used external antennas:

	C-ANT9103 Patch antenna (8x8) 6 dBi	5GHz Beamwidth 70°x70° ~33ft (10m)
	ANT2566P4W-R/S Patch antenna (4x4) 6 dBi	5GHz Beamwidth 110°x55° (120°x60°) ~33ft (10m)
	ANT2566D4M-R/S Patch antenna (4x4) 6 dBi	5GHz Beamwidth 55°x60° (60°x60°) ~33ft (10m)
	ANT2513P4M-N/S HD "Stadium" antenna 13 dBi	5GHz Beamwidth 31°x27° (30°x30°) ~66ft (20m)

This table lists the most commonly used integrated directional antennas:

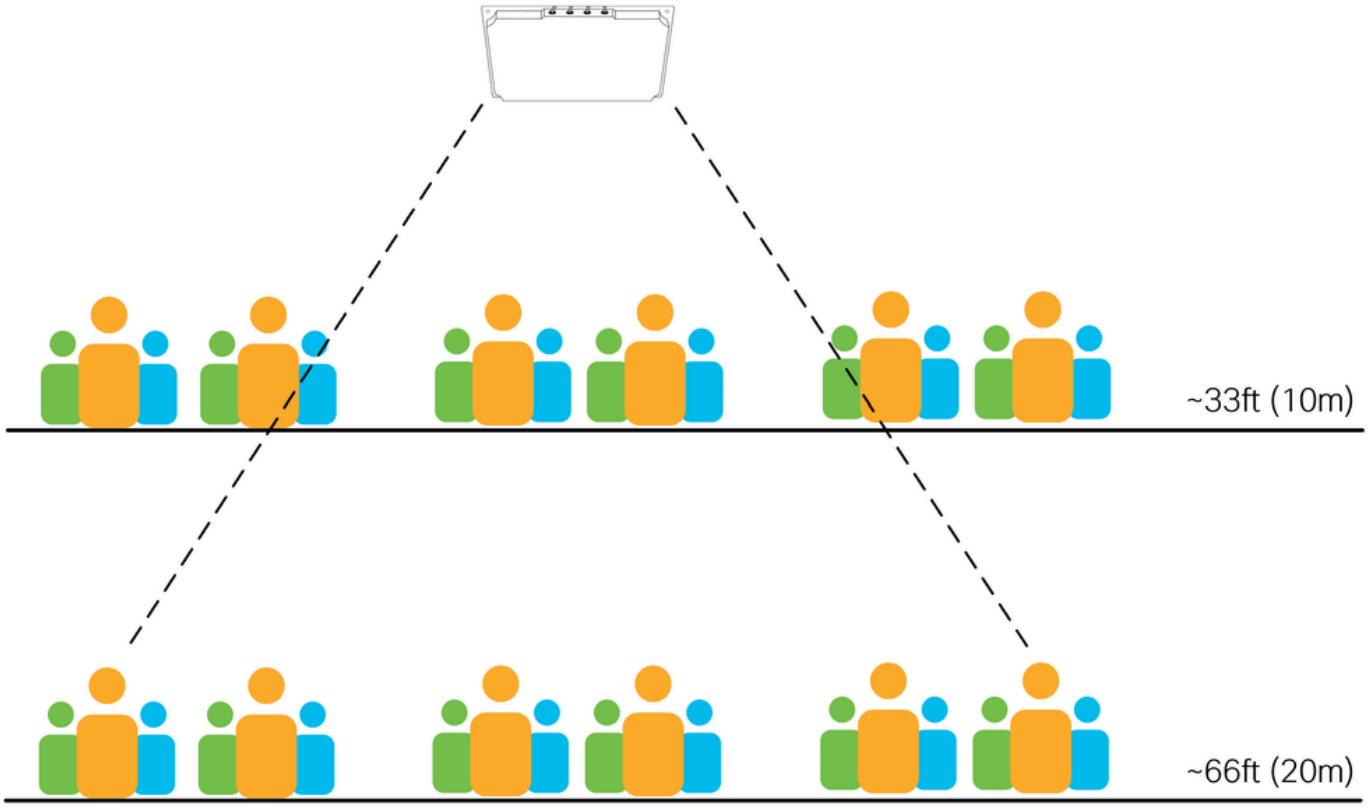
	CW9179F Integrated directional ~100ft (30m) *See data sheet for gain info.*	5GHz Narrow 35°x 35° Wide *See data sheet*	6GHz Narrow 35°x 35° Wide 70° x 35°
	C-ANT9104 Integrated directional ~66ft (20m)	5GHz Narrow 25°x 25° : 10dBi Wide 80°x 25° : 7dBi	
	CW9166D1 Integrated directional ~33ft (10m) *See data sheets for XOR info.*	6GHz 60°x 60° 8 dBi	5GHz 70°x 70° 6 dBi
	CW9176D1 Integrated directional ~33ft (10m) *See data sheets for XOR info.*	6GHz 70°x 70° 8 dBi	5GHz 70°x 70° 8 dBi

Main factors to consider when choosing an antenna are the antenna beamwidth and distance/height at which the antenna is mounted. The tables show beamwidth for commonly used antennas, with the numbers in brackets showing rounded (and easier to remember) values.

The suggested distances in the table are not hard rules, only guidelines based on experience. Radio waves travel at the speed of light and do not simply stop after reaching an arbitrary distance. The antennas all work beyond the suggested distance, however, performance drops as distance increases. Height of installation is key factor during planning.

The diagram below shows two possible mounting heights for the same antenna at ~33ft (10m) and ~66ft (20m) in a high-density area. Notice the number of clients that the antenna can *see* (and accept connections from) increases with distance. Maintaining smaller cell sizes becomes more challenging with larger distances.

The general rule is the higher the density of users, the more important it is to use the correct antenna for the given distance.



A stadium antenna

The CW9179F is well suited for covering high-density areas at high distances, see the *Cisco Wireless 9179F Access Point Deployment Guide* for of information.

Changes Over Time

Changes to the physical environment over time are common in almost all wireless installations (for example, movement of interior walls or remodeling of venue seating). Regular site visits and visual inspections have always been a recommended practice. For event networks there is the additional complexity of dealing with audio and lighting systems, and in many cases also other communication systems (such as 5G). All these systems are often installed in elevated locations above the users, sometimes resulting in contention for the same space. A good location for a wireless stadium antenna is often also a good location for a 5G antenna! Moreover, as these systems get upgraded over time, they can be relocated to locations where they obstruct and/or actively interfere with the wireless system. It is important to keep track of the other installations, and communicate with the teams installing them, to ensure that all the systems are installed in suitable locations without interfering with each other (physically or electromagnetically).

High Density and 6GHz

The cleaner and less congested 6GHz band offers better experience over 5GHz while having the added benefit of decreasing 5GHz airtime use as more client devices adopt 6GHz. Modern networks should be designed for both 5GHz and 6GHz coverage, including large public networks.
The Cisco product portfolio includes several 6GHz capable devices that can be used for high-density deployments; the 9179F for longer range and focused coverage, along with the 91x6D1 models for wider coverage at lower distances – the combination of these products allows for ample 6GHz design options across a variety of locations.
Designing 6GHz coverage generally follows the same methodology as 5GHz. Guidelines around the selection of data rates, channel widths and Rx-Sop values remain the same, although some additional consideration must be given to security configurations required for 6GHz as well as channel planning and the use of primary scanning (PSC) channels.

In the context of large public networks, these often have various large areas and require the use of a combination of antennas at various heights. It can be challenging to deploy a large public network end-to-end using only a 60°x60° antenna due to distance imitations. Therefore, it can also be challenging to provide end-to-end coverage at 6GHz using only the CW9166D1 for a large public network.

One possible approach is to make use of 5GHz as the primary coverage band, while using 6GHz only in specific areas to offload capable client devices to the cleaner 6GHz band. This type of approach makes use of 5GHz-only antennas in larger areas, while utilizing the 6GHz antennas where possible and where additional capacity is required.

Client roaming between 5GHz and 6GHz bands is discussed later in this document.

Regulatory

As with 5GHz, available power and channels for 6GHz differ significantly between regulatory domains. Notably, there is a large difference in available spectrum between FCC and ETSI domains, as well as strict guidelines around available Tx power for indoor and outdoor use, Low Power Indoor (LPI) and Standard Power (SP) respectively. With 6GHz, additional restrictions include client power limits, the use of external antennas and antenna down tilt, and (only in the US for now) the requirement for Automated Frequency Coordination (AFC) for SP deployments.

For more information on Wi-Fi 6E see *Wi-Fi 6E: The Next Great Chapter in Wi-Fi White Paper*.

Radio Resource Management

Radio Resource Management (RRM) is a set of algorithms responsible for control of radio operation. This guide references two key RRM algorithms, namely Dynamic Channel Assignment (DCA) and Transmit Power Control (TPC). RRM is an alternative to static channel and power configuration.

- DCA runs on a configurable schedule (default 10 minutes).
- TPC runs on an automatic schedule (default 10 minutes).

Cisco Event Driven RRM (ED-RRM) is a DCA option that allows for a channel change decision to be taken outside of the standard DCA schedule, usually in response to severe RF conditions. ED-RRM can change a channel immediately when excessive levels of interference are detected. In noisy and/or unstable environments enabling ED-RRM poses a risk of excessive channel changes, this is a potential negative impact to client devices.

The use of RRM is encouraged and generally preferred over static configuration - however, with certain caveats and exceptions.

- TPC must be limited to an appropriate range of values utilizing the TPC min/max setting, as needed, and always aligned to RF design.
    - Enable *TPC Channel Aware* in high-density environments.

- DCA cycle must be changed from the default setting of 10 minutes.
    - Do not use ED-RRM in busy,noisy and high-density environments.
    - Disable *Avoid Cisco AP Load*.
    - Rogue AP avoidance options like *Avoid Foreign AP Interference* can result in an unstable environment if there are many rogues. It is always better to remove the rogue than attempt to respond to it.
- RRM decisions can be impacted by APs/antennas that do not hear each other properly, as in the case of directional antennas that are pointing away from each other.
- Some antennas (C9104 for example) do not support RRM and always require static configuration.
- RRM does not fix poor RF design.

In all cases, RRM must be deployed with an understanding of the expected outcome, and tuned to operate within boundaries that are appropriate for the given RF environment. Subsequent sections of this document explore these points in more detail.

RF Configuration

Channels

In general, the more channels the better. In high-density deployments there can be orders of magnitude more APs and radios deployed than available channels, implying a large channel re-use ratio, and, along with that, higher levels of co-channel interference. All available channels must be used, and limiting the list of available channels is generally discouraged.

There can be instances where a specific (and separate) wireless system needs to co-exist in the same physical space, and dedicated channel(s) must be allocated to it, at the same time removing the allocated channel(s) from the DCA list of the primary system. These types of channel exclusions must be evaluated very carefully and used only when necessary. An example of this can be a point-to-point link operating in an open area adjacent to the primary network, or a press area inside a stadium. If more than one or two channels are being excluded from the DCA list, it is a cause for re-evaluation of the proposed solution. In some cases, such as very high-density stadiums, excluding even a single channel can sometimes not be a feasible option.

Dynamic Channel Assignment (DCA) can be used with WLC based RRM or AI-enhanced RRM.

The default DCA interval is 10 minutes which can result in frequent channel changes in unstable RF environments. The default DCA timer must be increased from the default 10 minutes in all cases, the specific DCA interval must be aligned with operating requirements for the network in question. An example configuration can be: DCA interval 4 hours, anchor time 8. This limits channel changes to once every 4 hours, starting at 8am.

As interferences are bound to happen, adapting to them every DCA cycle doesn't necessarily bring value as a lot of those interferences are temporary. A good technique is to use automatic DCA for the first few hours and freeze the algorithm and channel plan when you have something stable that you are happy with.

When the WLC is rebooted, DCA runs in aggressive mode for 100 minutes to find a suitable channel plan. It is a good idea to restart the process manually when significant changes are made to the RF design (for example adding or removing numerous APs or changing the channel width). To start this process manually use this command.

```
ap dot11 [24ghz | 5ghz | 6ghz] rrm dca restart
```



Note: Channel changes can be disruptive to client devices.

2.4 GHz

The 2.4GHz band has often been criticized. It only has three non-overlapping channels and many other technologies other than Wi-Fi use it, creating undesirable interferences. Some organizations insist on providing service on it, so what is a reasonable conclusion? It is a fact that the 2.4GHz band does not provide a satisfying experience for end users. Worse, by trying to provide service on 2.4GHz you affect other 2.4GHz technologies such as Bluetooth. In large venues or events, many people still expect their wireless headset to work when they place a call or their smart wearables to keep operating as usual. If your dense Wi-Fi operates at 2.4GHz you are impacting those devices who are not even using your 2.4GHz Wi-Fi.

One thing is certain, if you really must provide 2.4GHz Wi-Fi service, it is best to do that on a separate SSID (dedicate it to IoT devices or call it "legacy"). This means that dual-band devices do not connect to 2.4GHz involuntarily and only single-band 2.4GHz devices connect to it.

Cisco does not advise or support the use of 40MHz channels in 2.4GHz.

5 GHz

Typical deployment for high-density wireless. Use all available channels where possible.

Number of channels varies depending on regulatory domain. Consider the impact of radar at the specific location, use DFS channels (including TDWR channels) where possible.

20MHz channel width is *highly* recommended for all high-density deployments.

40MHz can be used on the same basis as 2.4GHz, that is only when (and where) absolutely needed.

Evaluate the need as well as real-world benefit of 40MHz channels in the specific environment. 40MHz channels require higher signal-to-noise ratio (SNR) to realize any possible improvement in throughput, if higher SNR is not possible then 40MHz channels serve no useful purpose. High-density networks prioritize average throughput for all users over potentially higher throughput for any single user. It is better to place more APs on 20MHz channels than having APs using 40MHz as the secondary channel is used only for data frames and therefore used much less efficiently than having two different radio cells, each operating on 20MHz (in terms of total capacity, not in terms of single client throughput).

The most critical detail is to keep the same channel width inside a given area. Indeed, if you configure different channel width, the clients have a tendency to stick to the APs with the largest channel width even if there is another AP closer at a stronger signal strength.

6GHz

The 6GHz band is not available yet in every country. Moreover, some devices have a 6GHz capable Wi-Fi adapter but require a BIOS update for it to get enabled for the specific country. The most popular way clients discover 6GHz radios right now is via the Reduced Neighbor Report (RNR) advertisement, this advertises the existence of 6GHz networks to clients using one of the other AP radios. This means that 6GHz must not operate alone without another radio on the same AP.

When configuring a dual-band SSID (5GHz & 6GHz) the 5GHz advertisement will include an RNR for the 6GHz SSID, allowing the client to quickly find the 6GHz SSID without having to scan the band.

Another approach is to configure a dummy (and/or hidden) SSID on 2.4GHz, the intention here is to advertise the 6GHz RNR using the 2.4GHz radio, without offering a 2.4GHz SSID. This approach is useful in situations where 6GHz capacity needs to be added without also adding 5GHz radios to the environment.

6GHz allows the use of larger channel bandwidths but this depends heavily on the number of channels available in the regulatory domain. With 24 6GHz channels available in Europe, it is not unreasonable to consider 40MHz channels to provide better maximum throughput compared to the 20MHz you are probably using in 5GHz. In the US, with nearly double the number of channels, using 40MHz is a no-brainer and even going for 80MHz is not unreasonable. Larger bandwidths must not be used in high-density events or venues.

The caveat with larger 6GHz channel widths is that it is not always possible to use the same channel width in 5GHz, and this will impact roaming behavior on some client devices. The most critical detail is to keep the same channel width inside a given area. Indeed, if you configure different channel width, the clients have a tendency to stick to the APs with the largest channel width even if there is another AP closer at a stronger signal strength, this also applies to roaming between the 5GHz and 6GHz bands.

Data rates

The data rate that a client negotiates with an AP is largely a function of the Signal-to-Noise Ratio (SNR) of that connection, and the opposite is also true, higher data rates *require* higher SNR. In fact, it is mostly SNR that determines the maximum possible link speed – but why is this important when configuring data rates? It is because some data rates have special meaning.

Classic OFDM (802.11a) data rates can be configured in one of three settings: Disabled, Supported, or Mandatory. The OFDM rates are (in Mbps): 6, 9, 12, 18, 24, 36, 48, 54, and the client and AP must both support a rate before it can be used.

Supported - the AP uses the rate

Mandatory - the AP uses the rate, and sends management traffic using this rate

Disabled - the AP does not use the rate, forcing the client to use another rate

---



Note: Mandatory rates are also referred to as Basic rates

The significance of the mandatory rate is that all management frames are sent using this rate, as well as broadcast and multicast frames. If there are multiple mandatory rates configured then management frames use the lowest configured mandatory rate, and broadcast and multicast use one of the mandatory rates (this point is configurable in the RF profile).

Management frames include beacons that must be heard by the client to be able to associate to the AP. Increasing the mandatory rate also increases the SNR requirement for that transmission, recall that higher data rates require higher SNR, and this typically means that the client needs to be closer to the AP to be able to decode the beacon and associate. Therefore, by manipulating the mandatory data rate we also manipulate the effective association range of the AP, forcing clients closer to the AP, or towards a potential roaming decision. Clients that are close to the AP use higher data rates, and higher data rates use less airtime - the intended effect is a more efficient cell. It is important to remember that increasing the data rate only affects the rate of transmission of certain frames, it does not affect the RF propagation of the antenna or the interference range. Good RF design practices are still needed to minimize co-channel interference and noise.

Conversely, leaving lower rates as mandatory typically means clients can association from a much further distance, useful in lower AP density scenarios, but with potential to cause havoc with roaming in higher density scenarios. Anyone that has tried to locate a rogue AP that is broadcasting at 6Mbps knows that you can detect the AP very far away from its physical location!

On the topic of broadcast and multicast, in some cases a second (higher) mandatory rate is configured to increase the rate of delivery of multicast traffic. This is seldom successful as multicast is never acknowledged and never retransmitted in case frames are lost. As some loss is inherent in all wireless systems, it is inevitable that some multicast frames are lost regardless of the configured rate. A better

approach to reliable multicast delivery are multicast-to-unicast conversion techniques that transmit the multicast as a unicast stream, this has the benefit of both higher data rates and reliable (acknowledged) delivery.

Using only a single mandatory rate is preferred, disable all rates below the mandatory rate, and leave all rates above the mandatory rate as supported. The specific rate to use depends on the use case, as already mentioned lower rates are useful in lower density and outdoor scenarios where distances between APs are larger. For high density and event networks low rates *must* be disabled.

If you are unsure where to start; use a 12Mbps mandatory rate for low density deployments, and 24Mbps for high-density deployments. Many large-scale events, stadiums, and even high-density enterprise office deployments have proven to work reliably with a 24Mbps mandatory rate setting. Appropriate testing is recommended for specific use cases where rates below 12Mbps or above 24Mbps are needed.



Note: It is best to leave all the 802.11n/ac/ax/be rates enabled (all the rates in the High Throughput section of the WLC GUI), there is seldom a need to disable any of these.

Transmit Power

Transmit power recommendations differ based on the deployment type. Here we differentiate indoor deployments using omnidirectional antennas, from those using directional antennas. Both types of antennas can exist in a large public network, although these would typically be covering different types of areas.

For omnidirectional deployments it is common to use automatic Transmit Power Control (TPC) with a statically configured minimum threshold, and in certain cases also a statically configured maximum threshold.



Note: TPC thresholds refer to radio transmit power and exclude antenna gain. Always ensure the antenna gain is configured correctly for the antenna model used, this is done automatically in the case of internal antennas and self-identifying antennas.

Example 1

TPC Min.: 5dBm, TPC Max.: Maximum (30dBm)

This would result in the TPC algorithm determining the transmit power automatically, but never going below the configured minimum threshold of 5dBm.

Example 2

TPC Min.: 2dBm, TPC Max.: 11 dBm

This would result in the TPC algorithm determining the transmit power automatically, but always staying between 2dBm and 11dBm.

A good approach is to create several RF profiles with different thresholds, for example low power (2-5dBm), medium power (5-11dbM), and high power (11-17dBM), then assigning omnidirectional APs to each RF profile as needed. The values of each RF profile can be adjusted to the intended use case and coverage area. This allows the RRM algorithms to operate dynamically while staying within pre-defined boundaries.

The approach for directional antennas is very similar, the only difference being the level of precision required. Directional antenna placement must be designed and verified during a pre-deployment RF survey, and the specific radio configuration values are typically an outcome of this process.

As an example, if a ceiling-mounted patch antenna is required to cover a certain area from a height of ~26ft (8m), the RF survey must determine the minimum Tx power required to achieve this intended coverage (this determines the minimum TPC value for the RF profile). Similarly, from the same RF survey we would understand the possible overlap required between this, and the next antenna, or even the point at which we want the coverage to end – this would provide the maximum TPC value for the RF profile.

RF profiles for directional antennas are typically configured either with the same minimum and maximum TPC values, or a narrow range of possible values (usually <= 3dBm).

RF profiles are preferred to ensure configuration consistency, static configuration of individual APs is not recommended. It is good practice to name RF profiles according to the coverage area, antenna type, and use case, for example RF-Auditorium-Patch-Ceiling.

The correct amount of Tx power is when the required SNR value is achieved by the weakest client in the intended coverage area, and no more than that. 30dBm is a great client SNR target value under real world conditions (that is, in a venue full of people).

CHD

Coverage Hole Detection (CHD) is a separate algorithm for identifying and remediating coverage holes. CHD is configured globally as well as per WLAN. A possible effect of CHD is the increase of Tx power to compensate for coverage holes (areas with clients consistently detected with poor signal), this effect is at radio level and affects all WLANs, even when triggered by a single WLAN configured for CHD.

Large public networks are typically configured to specific power levels using RF profiles, some can be in open areas with clients roaming in to and out of the areas, there is no need for an algorithm to dynamically adjust AP Tx power in response to these client events.

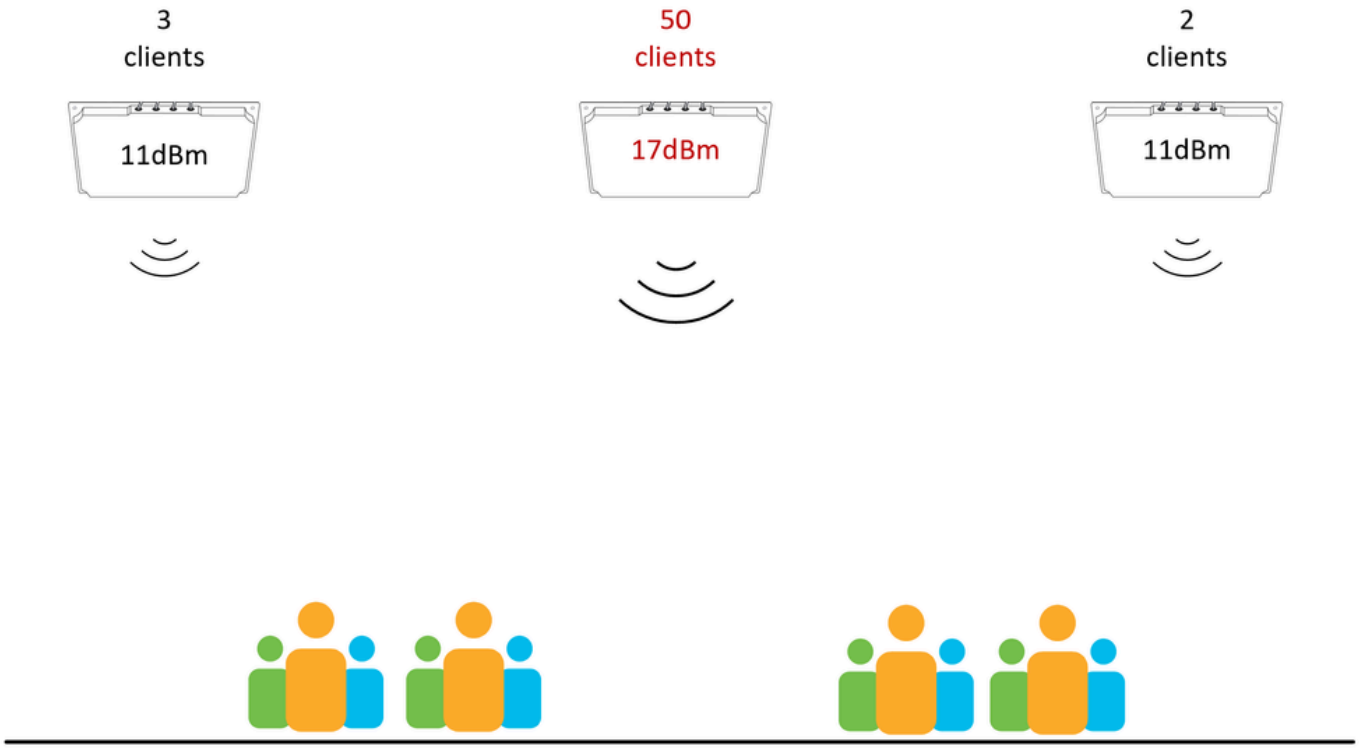
CHD must be disabled globally for large public networks.

Power Balance

Most client devices prefer a higher received signal when choosing which AP to associate to. Situations where an AP is configured with significantly higher Tx power compared to other surrounding APs must be avoided. APs operating at higher Tx power attract more clients, leading to uneven client distribution between APs (for example a single AP/radio is overloaded with clients while surrounding APs are underutilized). This situation is common in deployments with large coverage overlap from multiple antennas, and in cases where one AP has multiple antennas attached.

Stadium antennas such as the C9104 require particular care when selecting Tx power as the antenna beams overlap by design, please see relevant antenna deployment guides for more information on this.

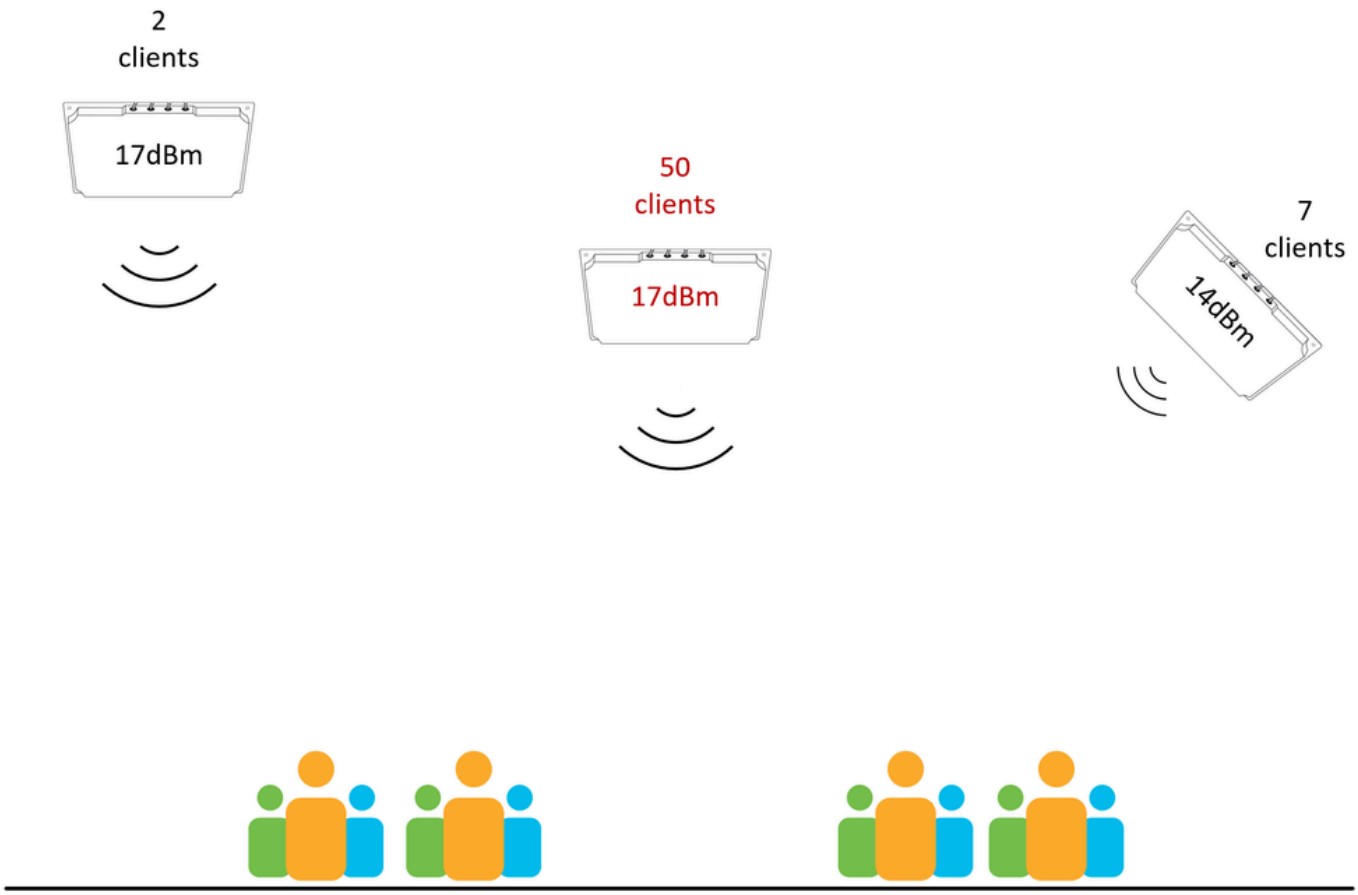
In the diagram below, the middle antenna is configured with a higher Tx power than the surrounding antennas. This configuration is likely to result in clients being ‘stuck’ to the middle antenna.



An AP with higher power than its neighbor APs attracts all the clients around

The next diagram shows a more complicated situation, not all antennas are at the same height, and not all antennas are using the same tilt/orientation. Achieving a balanced power is more complicated than simply configuring all radios with the same Tx power. In scenarios such as this, a post-deployment site survey can be required, this provides a view of the coverage from the client device point of view (on the ground). The survey data can then be used to balance the configuration for best coverage and client distribution.

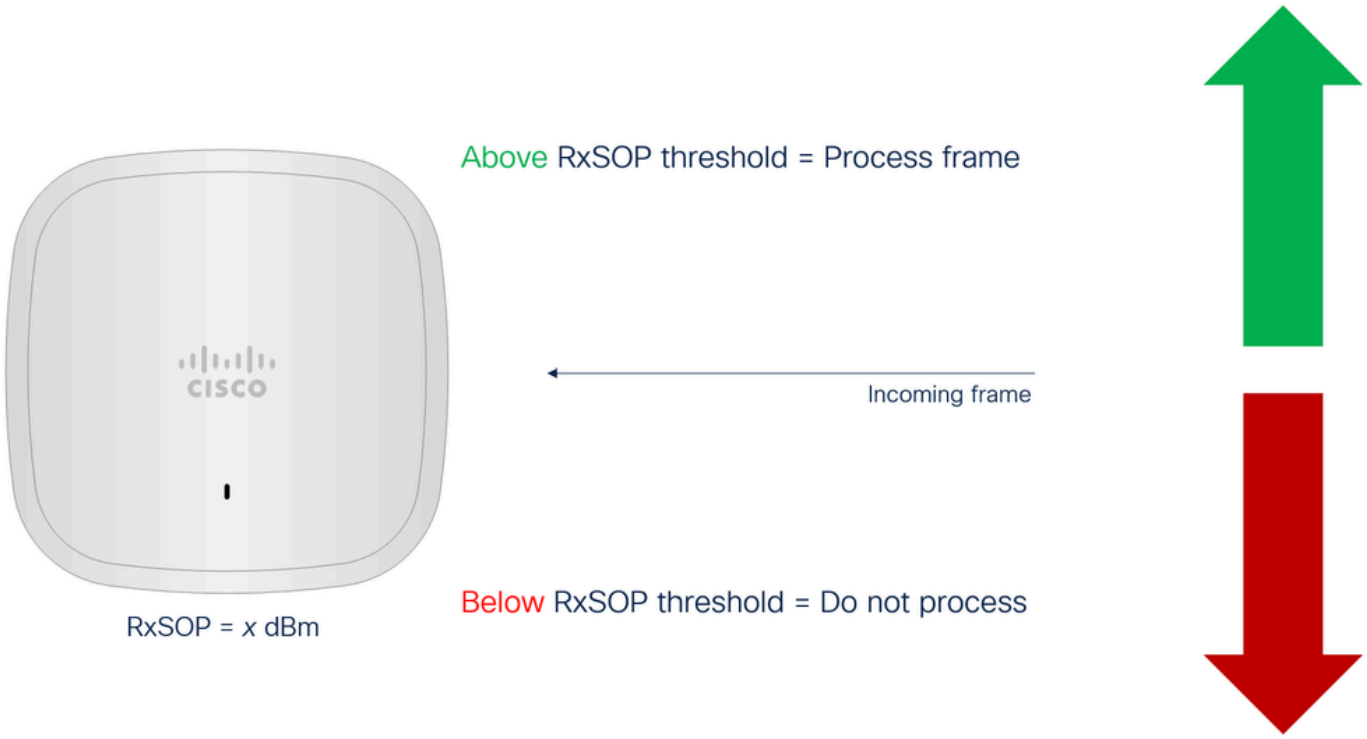
Designing uniform AP placement locations that avoid complicated situations like this is the best way to prevent challenging RF tuning scenarios (although sometimes there is no other choice!).



One AP is attracting all the clients despite Tx power being similar, but height and angles play a role

RxSOP

In contrast to mechanisms such as Tx power or data rates that affect the characteristics of the transmit cell, RxSOP (Receiver Start of Packet detection) aims to influence the size of the receive cell. In essence, RxSOP can the thought of as a noise threshold, in that it defines the received signal level below which the AP does not attempt to decode transmissions. Any transmissions arriving with a signal level weaker than the configured RxSOP threshold are not processed by the AP and are effectively treated as noise.



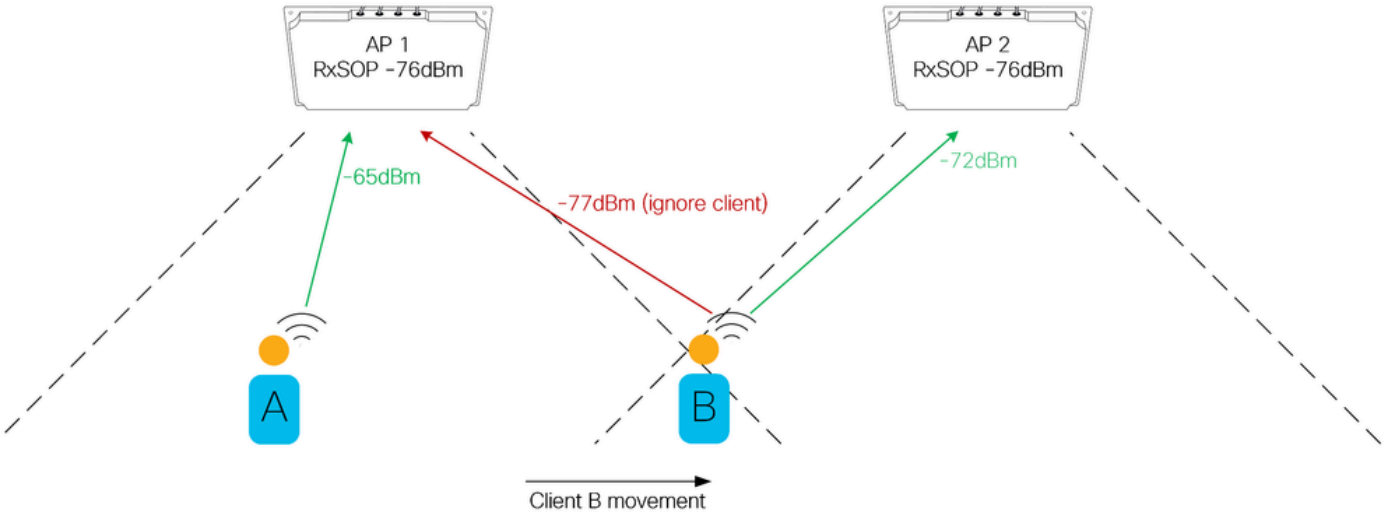
RxSOP concept explained

The significance of RxSOP

RxSOP has multiple uses. It can be used to improve the APs ability to transmit in noisy environments, to control the distribution of clients between antennas, as well as optimize for weaker and sticky clients.

In the case of noisy environments, recall that before transmitting an 802.11 frame the transmitting station (the AP in this case) first needs to assess the availability of the medium, part of this process is to first listen for transmissions that are already taking place. In dense Wi-Fi environments it is common for many APs to co-exist in a relatively limited space, often using the same channels. In such busy environments the AP can report channel utilization from those surrounding APs (including reflections) and delay its own transmission. By setting the appropriate RxSOP threshold the AP can ignore those weaker transmissions (reduction in perceived channel utilization) leading to more frequent transmit opportunity and improved performance. Environments where APs report significant channel utilization (for example > 10%) *without* any client load (for example an empty venue) are good candidates for RxSOP tuning.

For client optimization using RxSOP consider this diagram.



Client roaming affected by rxsop

In this example there are two APs/antennas with well-defined coverage areas. Client B is moving from the coverage area of AP1 into the coverage area of AP2. There is a crossover point at which AP2 hears the client better than AP1, but the client has not yet roamed to AP2. This is a good example of how setting the RxSOP threshold can enforce the boundary of the coverage area. Ensuring that clients are always connected to the closest AP improves performance by eliminating distant and/or weak client connections served at lower data rates. Configuring the RxSOP thresholds in this way requires a thorough understanding of where the expected coverage area of each AP begins and ends.

The dangers of RxSOP.

Setting the RxSOP threshold too aggressively results in coverage holes, as the AP is not decoding valid transmissions from valid client devices. This can have adverse consequences for the client as the AP does not respond; after all, if the client transmission was not *heard* there is no reason to respond. Tuning RxSOP thresholds must be done carefully, always ensuring the configured values do not exclude valid clients within the coverage area. Note that some clients can potentially no respond well to being ignored in this way, too aggressive RxSOP settings do not give the client a chance to roam naturally, effectively forcing the client to find another AP. A client that can decode a beacon from an AP assumes it is able to transmit to that AP, thus, the intention of RxSOP tuning is to match the size of the receive cell to the beacon range of the AP. Keep in mind that a (valid) client device does not always have direct line of sight to the AP, signal is often attenuated by users facing away from the antenna or carrying their devices in bags or pockets.

Configuring RxSOP

RxSOP is configured per RF Profile.

For each band there are preset thresholds (Low/Medium/High) at predefined dBm values. Those presets are different for each band (for example “Low” has a different value in dBm between 2.4GHz and 5GHz) The recommendation here is to always use custom values, even if the intended value is one of the available presets, this makes the configuration more readable. The “Auto” setting effectively disables RxSOP, in the sense that the AP tries to decode any packet it can as usual.

Setting	Value
Auto	Not configured
Low	-80dBm
Medium	-78dBm
High	-76dBm
Custom	-60dBm to -85dBm

RxSop settings table



Note: RxSOP changes do not require a radio reset and can be done on the fly.

Scaling the network

In general, using a device to the maximum of its documented capabilities is a bad idea. Data sheets report the truth, but the numbers mentioned can be in specific conditions of activity. Wireless controllers are tested and certified to support a certain number of clients and APs, and a certain throughput, but this does not assume that clients are roaming every second, that you can have configured extremely long unique ACLs for each client or enabled all the snooping features available. It is therefore important to consider all the aspects carefully to make sure the network scales during peak hours and to also keep a safety margin for future growth.

Number of APs

One of the first tasks in deploying any network is budgeting and ordering the right amount of equipment, and the largest variable factor is the number and type of access points and antennas. Wireless solutions must always be based on a radio frequency design, however (and unfortunately), very often this is the *second* step in the project lifecycle. In the case of simple indoor enterprise deployments there are numerous estimation techniques that can, to a reasonable level of certainty, predict how many APs can be required even before a wireless architect looks at the floorplans. Predictive models can also be very useful in this case.

For more challenging installations, such as industrial, outdoor, large public networks, or anywhere where external antennas are needed, simple estimation techniques are often inadequate. Some level of experience is required with previous similar installations to adequately estimate the type and amount of equipment needed. A site visit by a wireless architect is the bare minimum to get an understanding of the layout of a complex venue or facility.

This section provides guidelines on how to determine the **minimum** number of APs and antennas for the given deployment. Final quantities and specific mounting locations are **always** going to be determined through a requirement analysis and radio design process.

The initial bill of materials must be based on two factors: type of antennas, and quantity of antennas.

Type of antennas

There are no shortcuts here. The type of antenna is determined by the area that needs to be covered, and by the available mounting options in that area. It is not possible to determine this without an understanding of the physical space, this means a site visit is required by someone with an understanding of antennas and their coverage patterns.

Quantity of antennas

Quantity of equipment needed can be derived from an understanding of the expected amount of client connections.

Devices per Person

The number of human users can be determined by seating capacity of a venue, or number of tickets sold, or expected number of visitors based on historical statistics. Each human user can carry multiple devices and it is common to assume more than one device per user, although the ability of a human user to actively use multiple devices at the same time is questionable. The number of visitors that actively connect to the network also depends on the type of event and/or deployment.

Example 1: It is normal that an 80,000-seat stadium does not have 80,000 connected devices, this percentage is usually significantly lower. Connected user ratios of 20% are not uncommon during sports events, this means that for the 80,000-seat stadium example, the expected number of connected devices can be 16,000 (80,000 x 20% = 16,000). This number also depends on the onboarding mechanism used, if the user is required to perform some action (such as click a web portal) then numbers are lower than when device onboarding is automatic. Automatic onboarding can be as simple as a PSK that has been remembered from a previous event, or something more advanced like the use of OpenRoaming that onboards devices without user interaction. OpenRoaming networks can drive the user connection ratios well above 50%, which can have a significant impact on capacity planning.

Example 2: It is reasonable to expect that a technology conference has a high user connection ratio. Conference attendees spend longer connected to the network and expect to be able to access their email and perform daily tasks throughout the day. It is also more likely that this type of user connects more than one device to the network – although they will still likely only use one device at time. For technology conferences the assumption is that 100% of visitors connect to the network, this number can be lower for depending on the conference type.

In both examples, the key is to understand the expected number of connected devices and there is no single solution for every large public network. In either case, an antenna is attached to a radio, and it is client devices (not human users) that connect to that radio. Therefore, client devices per radio is a usable metric.

Devices per Radio

Cisco APs have a maximum client count of 200 connected devices per radio for Wi-Fi 6 APs and 400 devices per radio for Wi-Fi 6E APs. However, it is not advisable to design for maximum client count. For planning purposes, it is recommended to keep the client count per radio well below 50% of the maximum

AP capacity. Additionally, the number of radios depends on the type of AP and antenna used, the section on single vs dual 5GHz explores this in more detail.

At this stage it is a good idea to break up the network into distinct areas, with expected device counts per area. Recall, this section aims to estimate a **minimum** number of APs and antennas.

Consider an example of three distinct coverage areas, the expected client count is provided for each area, and a (healthy) value of 75 clients per radio is used to estimate the number of radios required.

Area	Expected Devices	Devices / Radio	Radios
Area 1	1000	75	14
Area 2	2000	75	27
Area 3	2500	75	34
Total			75

Expected radios/client count per area

These initial numbers now need to be combined with the understanding of what types of APs and antennas are deployed in each area, and if single or dual 5GHz is used. 6GHz calculations follow the same logic as 5GHz. 2.4GHz is not taken into consideration in this example.

Let's assume that each of the three areas uses a combination of 2566P patch antenna and the 9104 stadium antenna, with a combination of singe and dual 5GHz – this scenario is used for illustration purposes.

Area	Total Radios	2566P (Dual 5GHz)	2566P (Single 5GHz)	9104 (Dual 5GHz)
Area 1	14	0	6	4
Area 2	27	6	3	6
Area 3	34	7	0	10
Total Antennas		26	9	20
Total APs		13	9	0 (integrated)

Antennas per area

Each area lists the type of antennas and APs needed. Note that in the case of dual 5GHz the ratio is two antennas to one AP.

This section demonstrates an approach to estimate an initial number of antennas and APs needed for a

deployment. The estimate requires an understanding of the physical areas, possible mounting options in each area, the type of antennas to be used in each area, and number of client devices expected.

Each deployment is different and additional equipment is often needed to cover specific or challenging areas, this type of estimate only considers client capacity (not coverage) and serves to outline the scale of the investment needed. Final AP/antenna placement locations and equipment totals are **always** subject to a thorough understanding of the use-case and on-site verification by an experienced wireless professional.

Expected Throughput

Each wireless channel can offer an amount of available capacity which is typically translated to throughput. This capacity is shared between all devices connected to the radio, meaning that performance for each user drops as more user connections are added to the radio. This drop in performance is *not linear* and is also dependent on the exact mix of clients connected.

Client capabilities differ between devices depending on the client chipset and the number of spatial streams the client supports. Maximum client data rates for each number of supported spatial streams are listed in the table below. Note that most client devicess support a maximum of 2 spatial streams.

Client Capability	20MHz channel Wi-Fi 5 (802.11ac)	20MHz channel Wi-Fi 6 (802.11ax)
1 Spatial Stream(s)	86.7Mbps	121.9Mbps
2 Spatial Stream(s)	173.3Mbps	243.8Mbps
3 Spatial Stream(s)	288.9Mbps	365.6Mbps
4 Spatial Stream(s)	346.7Mbps	487.5Mbps

Expected maximum real throughput for each client type

The rates listed are theoretical maximum MCS (Modulation and Coding Scheme) rates derived from the 802.11 standard and assume a signal-to-noise ratio (SNR) >30dBm. The main design goal of well-performing wireless networks is to achieve this level of SNR for all clients in all locations, this is however rarely the case. Wireless networks are dynamic in nature and use unlicensed frequencies, various uncontrolled interferences have an impact on client SNR, in addition to client capabilities.

Even in cases where the required level of SNR is achieved, the rates listed previously do not consider protocol overhead, therefore, do not map directly to real world throughput (as measured by various speed test tools). Real world throughout is always lower than the MCS rate.

For all wireless networks (including large public networks), client throughput *always* depends on:

- Capabilities of the client.
- Signal to noise ratio of the client at that specific point in time.

- Number of other clients connected at that specific point in time.
- Capabilities of other clients at that specific point in time.
- Activity of other clients at that specific point in time.
- Interference at that specific point in time.

Based on the variability of these factors it is not possible to guarantee a minimum per-client throughout for wireless networks, regardless of the equipment vendor.

For more information, please reference the *Validate Wi-Fi Throughput: Testing and Monitoring Guide.*

WLC Platform

Choosing your WLC platform can seem easy. The first thing you can think about is to look at the estimated AP count and client count you aim to manage. The data sheet for each WLC platform contains all the maximum supported *objects* on the platform: ACLs, client count, site tags and so on. Those are literal maximum numbers and often there is a hard enforcement. You cannot join 6001 APs to a 9800-80 that supports only 6000 APs, for example. But is it wise to aim for the maximum everywhere?

Cisco wireless controllers are tested to be able to reach those maximums, but they cannot necessarily reach all documented maximums in all conditions at the same time. Let’s take the example of throughput, a 9800-80 can reach up to 80Gbps of client data forwarding, but this is in the case where each client packet is the maximum and optimum size of 1500 bytes. With a mix of packet sizes, the effective maximum throughput is lower. If you enable DTLS encryption the throughput gets further reduced, and the same for Application Visibility. It is optimistic to expect more than 40Gbps out of a 9800-80 in realistic conditions on a large network with many features enabled. Since this varies greatly depending on the features in use and the type of network activity, the only way to get an actual idea of the capacity is to measure the datapath utilization using this command. Focus on the *load* metric, which is a percentage of the maximum throughput the controller can forward.

```
WLC#show platform hardware chassis active qfp datapath utilization summary

  CPP 0:                      5 secs        1 min        5 min       60 min

Input:     Total (pps)            9            5            5            8

                  (bps)        17776         7632         9024        10568

Output:    Total (pps)            5            3            3            6

                  (bps)        11136        11640        11440        41448

Processing: Load (pct)            0            0            0            0


WLC#
```

In a similar manner, the 9800-80 can perfectly handle 6000 APs with regular activity. However, 6000 APs in a public venue such as a stadium or an airport do not count as regular activity. Considering the amount of

client roaming and ambient probing, large public networks at maximum scale can cause increased CPU utilization on a single WLC. If you add monitoring and SNMP traps to be sent each time clients move around, the load can quickly become too much. One of the key specifics of a large public venue or large event is that there are significantly more client onboarding events as people move around and constantly associate/disassociate, so this causes extra pressure on the CPU and control plane.

Numerous deployments have shown that a single (HA) pair of 9800-80 wireless controllers can handle a large stadium deployment with well over 1000 APs. It is also common to distribute the APs over two or more controller pairs for critical events where uptime and availability are primary concerns. When large networks are distributed over multiple WLCs there is the additional complexity of inter-controller roaming, client roaming must be considered carefully in confined spaces such as a stadium bowl.

See also the *Site Tag* section in this document.

WLC High-Availability

It is advised to use a High-Availability Stateful Switch Over (HA SSO) pair, this provides hardware redundancy but also protects against software failure. Using HA SSO, a software crash on one device is transparent to the end users as the secondary WLC takes over seamlessly. Another advantage of an HA SSO pair is the hitless upgrades offered by the In-Service Software Upgrade (ISSU) feature.

If the network is large enough, it is also advised to use an extra controller (N+1). It can serve several purposes that the HA SSO cannot fulfill. You can test a new software version on this WLC before upgrading the production pair (and migrate only a few test APs to it to test a specific section of the network). Some rare conditions can affect both WLCs in an HA pair (when the problem is replicated to the standby) and here the N+1 allows to have a safe WLC in an active-active scenario where you could progressively migrate APs to and from. It could also serve you as a provisioning controller to configure new APs.

The 9800-CLs are very scalable and powerful. It is to be noted that they have a much smaller data forwarding capacity (From 2 Gbps to 4 Gbps for the SR-IOV image) which tends to restrict them to FlexConnect local switching scenarios (and possibly a small number of APs in central switching). They can however be helpful as N+1 devices when you need extra controllers during a maintenance window or when troubleshooting a problem.

External Systems

While this document focuses mainly on the wireless component of large event networks, there are also numerous supporting systems that require consideration during the scaling and design phase, some of these are discussed here.

Core Network

Large wireless networks are typically deployed in central switching mode and with large subnets. This implies that a very large number of client MAC address and ARP entries are pushed out to the adjacent wired infrastructure. It is critical that the adjacent systems dedicated to the various L2 and L3 functions

possess the adequate resources to handle this load. In the case of L2 switches a common configuration is the adjustment of the Switch Device Manager (SDM) template, which is responsible for the allocation of system resources, balancing between L2 and L3 features depending on the function of the device within the network. It is important to ensure that core L2 devices can support the number of MAC address entries expected.

Gateway NAT

The most common use-case of public networks is to provide Internet access to visitors. Somewhere along the data path there must be a device responsible for NAT/PAT translation. Internet gateways must possess the required hardware resources and IP pool configuration to handle the load. Remember that a single wireless client device can be responsible for numerous NAT/PAT translations.

DNS/DHCP

These two systems are key to ensuring a good client experience. Both DNS and DHCP services require not only the appropriate scaling to handle the load, but also consideration with regards to placement within the network. Fast and responsive systems, placed at the same location as the WLC ensures the best experience and avoid long client onboarding times.

AAA/Web portal

No one likes a slow web page, choosing an appropriate and well-scaled system for external web authentication is important for a good client onboarding experience. Similarly for AAA, RADIUS authentication servers must be able to cope with the demands of the wireless system. Keep in mind that in some cases the load can spike during key moments, for example half-time during a football match, which can generate high authentication load in a small amount of time. Scaling the system for adequate *concurrent* load is key. Specific care must be taken when using features such as AAA accounting. Avoid time-based accounting at all costs and if you use accounting try to disable interim accounting. Another important item to consider is the use of load-balancers, here session-pining mechanisms must be used to ensure complete authentication flows. Make sure to keep the RADIUS timeout at 5 seconds or higher.

If using an 802.1X SSID with a large client count (for example with OpenRoaming), make sure to enable 802.11r Fast Transition (FT), otherwise clients can cause an authentication storm every time they roam around.

DNS/DHCP

A few recommendations for DHCP:

- Ensure the DHCP pool is at least three times the number of clients expected. IPs stay assigned for some time even after the client disconnected, so depending on the dwell time of guests this can consume more IP addresses. Try to match the lease time to the expected duration of the user’s visit to the venue, there is no point in allocating an IP address for a week if a typical visit duration is two hours, this helps to cycle out stale leases.
- Using a single large subnet for clients is recommended, the WLC has a proxy ARP feature and

doesn’t forward broadcasts by default (other than DHCP). Using a large (for example /16) client subnet for your clients does not represent a problem. A single large VLAN is simpler compared to a VLAN group with many VLANs. Configuring many smaller subnets (for example /24) and VLANs groups does not influence the broadcast domain and only results in a more complicated configuration, resulting in issues like dirty VLANs and having to keep track of various DHCP pools that can not get used evenly. This recommendation applies to Local mode APs and centrally switched traffic. In case of FlexConnect with local switching, it is not supported to put more than 100 (or 300 after 17.9) APs in the same subnet as they are suffering from a lot of broadcast transmissions. Flexconnect requires APs to serve the same VLAN in order to support seamless roaming and therefore it is not advised to support large subnets in this scenario.

- Keep DHCP in bridging mode on the wireless controller with the DHCP relay functionality handled by the Layer 3 gateway of the subnet. This allows for maximum efficiency and simplicity. The idea is to not have the wireless controller involved in the DHCP process at all.
- Use DHCP Required on any public WLAN, regardless of authentication method. Although this can trigger a small percentage of failed client associations, it could prevent significant security issues either by clients trying to assign themselves static IP addresses or by clients misbehaving and trying to reuse a previous IP address without permission.

Operating the Network

The Right Configuration

It is tempting to enable a lot of options to benefit from all the latest features of modern Wi-Fi. However, certain features work great in small environments but have a huge impact in large and dense environments. Similarly, certain features can pose compatibility problems. Even though Cisco equipment respects all the standards and offers compatibility with a wide variety of tested clients, the world is filled with unique client devices that sometimes have driver software versions with bugs or incompatibility with certain features.

Depending on the level of control you have on the clients, you must be conservative. For example, if you deploy the Wi-Fi for the large annual gathering of your company, you know that most clients are company devices and you can plan the feature set to enable accordingly. On the other hand, if you operate an airport Wi-Fi, your guest satisfaction level directly relates to their ability to connect to your network, and you have no control whatsoever on the client devices people can be using.

SSIDs

How many SSIDs?

The recommendation has always been to use as few SSIDs as possible. This is exacerbated in high density networks as the possibility of having several APs on the same channel is almost guaranteed. Typically, many deployments use too many SSIDs, acknowledge they have too many SSIDs, but declare that they cannot use less. You must perform a business and technical study for each SSID to understand the similarities between SSIDs and options for collapsing multiple SSIDs into one.

Let’s go over a few security/SSID types and their use.

WPA2/3 Personal

A pre-shared key SSID is immensely popular due to its simplicity. You can either print the key somewhere on badges or on paper or signs or communicate it somehow to visitors. Sometimes a pre-shared key SSID is preferred even for a guest SSID (provided the key is well-known by all attendees). It can help preventing DHCP pool exhaustion due to the deliberate nature of the connection. Devices passing by do not connect automatically to the network, therefore are unable to consume an IP address from the DHCP pool.

WPA2 PSK does not provide privacy as traffic can easily be decrypted since everyone uses the same key. On the contrary, WPA3 SAE does provide privacy, and even if everyone has the master key it is not possible to derive the encryption key used by other clients.

WPA3 SAE is the better choice for security and many smartphones, laptops and operating systems support it. Some IoT devices or smart wearables can still have limited support and older clients in general are susceptible to issues if they did not receive recent drivers or firmware updates.

It can be tempting to consider a Transition mode WPA2 PSK-WPA3 SAE SSID to simplify things, but this has been shown in the field to cause some compatibility issues. Poorly programmed clients do not expect two types of shared key methods on the same SSID. If you want to offer both WPA2 and WPA3 options, it is advised to configure separate SSIDs.

Wi-Fi 7 requires two new AKMs (SAE-EXT and SAE-EXT-FT) to be configured on the WPA3 SSID as well as GCMP256 encryption which reinforces the need to make a choice (if you want to keep WPA2 compatibility, it is potentially preferable to forget about Wi-Fi 7 capability). Not all AP models nor clients support GCMP256 at this time.

WPA2/3 Enterprise

WPA3 Enterprise (using AES 128-bit encryption) is technically the same security method (at least as advertised in the SSID beacons) as WPA2 Enterprise, which provides for maximum compatibility.

For 802.1X, a transition mode SSID is advised as compatibility problems are not seen with recent devices (problems were reported with Android 8 or old Apple IOS versions). IOS XE 17.12 and later releases allow to have a single Transition Enterprise SSID where only WPA3 is used and advertised on 6GHz while WPA2 is offered as an option on the 5GHz band. We advise to enable WPA3 on Enterprise SSIDs as soon as possible.

WPA Enterprise SSIDs can be used for key users for which there is an identity provider database that allows to return AAA parameters (such as VLANs or ACLs) depending on the user identity. Such types of SSIDs can include eduroam or OpenRoaming which combine the benefits of guest SSIDs (by allowing visitors to connect easily without entering any credentials) with the security of a corporate SSID. They greatly reduce the complexity of onboarding typically associated with 802.1X as clients do not have to do anything to join the eduroam or OpenRoaming SSID, provided they have a profile on their phone (which can be easily provided through an event app)

Wi-Fi 7 support can require (with the strictest clients) the AP to also offer GCMP256 encryption which requires IOS-XE 17.18.3 and later software.

Guest SSIDs

A guest SSID is often synonymous with open authentication. You can add a web portal (or not) behind it (depending on the desired friendliness or local requirements) in its various forms: external, local, or central web authentication, but the concept stays the same. When using a guest portal, scalability can quickly become a problem in large environments. Check the *Configuring for Scalability* section for more information on this.

6GHz operations require your guest SSID to use Enhanced Open rather than just Open. This still allows anyone to connect but provides privacy (a better privacy than WPA2-PSK even!) and encryption, all without providing any key or credentials when connecting on the SSID. Main smartphones vendors and operating systems now support Enhanced Open, but the support is not yet widespread in the wireless client base. Enhanced Open transition mode is forbidden with Wi-Fi 7 and 6GHz operations. Therefore, the right approach is probably to run the Enhanced Open SSID on the 6GHz band only and keep the fully Open SSID on the 5GHz band for compatibility. You need to give the SSIDs a different name since they have different security methods. This adds an SSID in the client scan list but does not add an extra SSID over the air since they are present on different bands.

In large events and venues, it is often advised to configure a PSK on the Guest SSID rather than leaving it purely open (Enhance Open Transition mode would be better but that creates two SSIDs and client compatibility still must be extensively proven). Although this makes the onboarding a bit more complicated (you must print the PSK on people's badges or tickets or advertise it somehow), it avoids casual clients connecting to the network automatically without the end user having any intention to use the network. More and more mobile operating system vendors also de-prioritize open networks and show a security warning. In other situations, you can want a maximum number of passersby to connect and therefore open is the better choice.

Conclusion on the number of SSIDs

There cannot be a satisfactory answer to the question of how many SSIDs do you have to stick to. The effect depends on the minimum configured data rate, the number of SSIDs and the number of APs broadcasting on the same channel. At one large Cisco event, the wireless infrastructure used 5 SSIDs: the main WPA2 PSK, a WPA 3 SAE SSID for security and 6GHz coverage, an enterprise Eduroam SSID for ease of access for educational attendees, an OpenRoaming SSID to securely welcome anyone who configured Wi-Fi from the event app and a separate 802.1X SSID for the staff and admin network access. This was nearly too much already, but the effect stayed reasonable thanks to the large number of channels available, and the directional antennas used to reduce channel overlap as much as possible.

The Legacy SSID versus Main SSID concepts

For a certain period, it was advised to restrict 2.4GHz service to a "Legacy" separate SSID only advertised in 2.4GHz. This is getting less popular as people stop providing 2.4GHz service altogether. However, the idea can and must persist but with other concepts. You want to roll out WPA3 SAE, but transition mode is

giving you compatibility issues with your clients? Have a WPA2 “Legacy” SSID and a main WPA3 SAE SSID. By naming the least performing SSID “legacy” it does not attract clients and you are able to see easily how many clients still face compatibility issues with your main SSID and require this legacy one.

But why stop there? You heard rumors that 802.11v caused issues with some older clients or that some client drivers do not like seeing Device Analytics enabled on the SSID? Enable all those handy features on your advanced main SSID and leave them off on your legacy/compatibility SSID. This allows you to test the roll out of new features on your main SSID while still providing a maximum compatibility SSID for clients to fall back to. This system only works this way. If you do the opposite name your compatibility-driven SSID as your main and name your advanced SSID with something like “<name>-WPA3”, you notice people sticking to the old SSID they were used to, and adoption stay small for many years on your “new” SSID. Rolling out new settings or features then have inconclusive results due to the lower number of clients connecting to it.

SSID features

- It is best to keep Aironet Extensions disabled. Those are particularly useful for site surveys and WGB operations, but sometimes cause issues with some legacy clients. Aironet IE also advertises the AP hostname which is unwanted in security-conscious deployments.
- CCKM is a deprecated protocol (in favor of FT) and must be disabled.
- At this time, it is best to use AES-128 encryption, even in WPA3 due to low client support of higher encryptions (unless you can afford a specific more secure and restrictive SSID)
- Coverage Hole Detection is best disabled (for all SSIDs). Large deployments typically use directional antennas, requiring a thorough site-survey. The power levels of each antenna would be an outcome of the RF design process and typically configured to specific levels.
- Adaptive FT must be disabled as some clients can have issues when FT is not fully advertised but present in some attributes. Either fully disable FT (for maximum compatibility) or go with FT+802.1X which most clients (unless they are old or more IoT oriented) do support. When configuring FT+802.1X, even non-FT clients are allowed to join the SSID. The only possible problem is with some clients who would not tolerate seeing two security options on the same SSID.
- Disable 802.11ac MU-MIMO. It adds complexity and has very low benefit in 802.11ac.
- Disable BSS Target Wake Time. It has low adoption on the client side currently.
- Disable aggressive load-balancing and Band Select. Band Select is not needed if you do not advertise the SSID in 2.4GHz (or if it is on a dedicated SSID) and aggressive load-balancing delays client association by rejecting the client a couple of times before finally accepting it if it insists on connecting to a loaded AP. You have loaded APs anyway in a busy environment and this is negative for the client experience.
- Disable Fastlane+.
- Disable Universal Admin, this feature was for 3700 AP and only in the -UX domain. Leaving it on leaves open an unnecessary attack vector.
- Keep Opportunistic Key Caching (OKC) enabled. It serves as fast roaming mechanism for clients not supporting FT.
- Keep WMM allowed. Disabling it would take your network back to the 802.11g era and requiring it doesn’t bring any advantage on the 9800 platform.
- Enable IP Source Guard.
- Disable RADIUS profiling. In a very busy environment, this can send excessive RADIUS accounting messages (whenever the clients do DHCP or send HTTP packets) and has a very real potential of overloading your RADIUS server.
- Avoid using hidden SSIDs. This does not serve any security purpose, the SSID name can still be discovered easily with simple applications or by taking a sniffer capture. Hiding the SSID slows down all client roaming as they do not benefit from passive beacon scanning anymore and must rely on

active scanning to get neighboring AP information.
- Try not using more than four WLANs per radio as it has a significant impact RF utilization. It is not a hard limit, using five WLANs can work but be very conscious of the wasted airtime by using more and more WLANs.
- 802.11v and 802.11k are standards that are more and more supported by popular client types. They typically do not pose a problem with regards to client connection. The benefits they bring depend a lot on how clients utilize those protocols and can sometimes (in the case of 802.11k) cause a slightly higher CPU usage. You can keep them out of your IoT or Legacy SSID but they must be enabled if possible on your production SSID.

Site Tag

Site tags are a configuration item which allows to group access points who share the same FlexConnect settings as well as AP join profile settings (such as credentials, SSH details and country code). Why are site tags important? Site tags also define how APs are handled by the WNCD process inside the Catalyst 9800. Let’s give a few examples to illustrate:

- If you configure four site tags on a 9800-80 which has eight WNCD processes, each site tag gets assigned to a different WNCD process (running each on a separate CPU core) and four WNCD processes do nothing. This means you are not utilizing all the CPUs of your 9800-80 and it would not be recommended to load it with the maximum 6000 supported APs.

Site tag 1	Site tag 2	Site tag 3	Site tag 4	-	-	-	-
WNCD 1	WNCD 2	WNCD 3	WNCD 4	WNCD 5	WNCD 6	WNCD 7	WNCD 8
CPU	CPU	CPU	CPU	CPU	CPU	CPU	CPU

First example of site tag balancing

- If you configure 10 site tags on a 9800-80 that has eight WNCD processes, two WNCD processes take care of two site tags each, while the remaining six handle one site tag each.

Site tag 1 Site tag 9	Site tag 2 Site tag 10	Site tag 3	Site tag 4	Site tag 5	Site tag 6	Site tag 7	Site tag 8
WNCD 1	WNCD 2	WNCD 3	WNCD 4	WNCD 5	WNCD 6	WNCD 7	WNCD 8
CPU	CPU	CPU	CPU	CPU	CPU	CPU	CPU

Second example of site tag balancing

For geographically large deployments with many sites and many site tags, the number of site tags is recommended to be a multiple of the number of WNCD processes on the platform you are using.

However, for event networks that are typically under one roof, or multiple buildings at the same venue, the recommendation is to match the number of site tags to the exact number of WNCDs on the given platform.

The end goal is that each WNCD process (and therefore each CPU core allocated to wireless tasks) handles a roughly similar number of client roam events so that the load is balanced across all CPU cores.

Platform type	Number of WNCD processes
9800-CL small OVA	1
9800-CL medium OVA	3
9800-CL large OVA	7
9800-L	1
9800-40/CW9800-M	5
9800-80/CW9800-H	8

Number of WNCD process for each platform type

At the core, what really matters is to group APs that are in the same physical neighborhood into the same site tag, so that the frequent client roaming events between these APs stay in the same CPU process. This means that even if you have a single large venue, it is recommended to divide the venue into several site tags (as many as you have WNCD processes handling the venue) and group APs as logically as possible into these to form logical RF neighborhood groups that are also evenly distributed among site tags.

Starting IOS XE 17.12, a load balancing algorithm can be enabled so that the WLC groups the APs based on their RF proximity. This takes the burden out of your hands and creates a balanced spread of the APs across WNCD process. This can be helpful if you cannot easily draw groups of neighbor APs to be placed in the correct amount of site tags. One specificity of this algorithm is that it assigns APs to WNCD process regardless of their site tag assignment, this means it does not change the site tag assignment of the AP. You can then assign site tags purely basic on a configuration logic and let the algorithm balance the APs across CPUs in the most optimal manner.

The RF based Automatic AP Load Balancing feature is documented in *Cisco Catalyst 9800 Series Wireless Controller Software Configuration Guide, Cisco IOS XE Dublin 17.12.x.*

CPU usage of WNCD processes must be monitored during large events. If one or more WNCD processes show high utilization, it can be that the WNCD is handling too many APs or clients, or that the APs or clients it handles are busier than the average (if all of them constantly roam such as in an airport for example).

Policy Profile

- Enable ARP and Duplicate Address Detection (DAD) Proxy, this allows the WLC to reply on behalf of wireless clients when a device is trying to learn the MAC address of a wireless device. This also saves wireless client batteries.
- Do not enable WGB features unless needed.
- Enable DHCP required to avoid clients with static IP addresses.
- Keep idle-timeout short (300 seconds). Some administrators make it long to avoid clients having to re-authenticate, but long idle-timeout results in ghost client entries (affecting reporting) as the client count becomes delayed from real-time. It is best to keep the idle -timeout lower than the group key

rotation timer to avoid accounting floods when the clients are deleted. The group key rotation interval can be configured in the web UI under **Configuration > Security > Advanced EAP** as "EAP-Broadcast Key Interval"
- Make the session timeout 86400 seconds to avoid unnecessary disconnections and re-authentications.

AP Join Profile

- Ensure TCP adjust MSS is enabled.
- Enable Trust DSCP upstream. Many wireless clients do not do 802.11e WMM UP tagging unfortunately, trusting the DSCP field is a sure way to provide the right priority to voice applications.
- Enable Syslog for your access points. Configuring a Syslog server IP makes the APs unicast their console logs to it. It is not only useful to troubleshoot APs, but it is also better for the network than the default setting which makes APs broadcast their Syslog in the local VLAN. AP logging can generate significant message load, even in cases where AP Syslog is *not* monitored it is still a good idea to limit the number of events by setting the appropriate message severity, and/or configuring a dummy Syslog IP address (for example 0.0.0.0) to prevent messages from being broadcasted.
- Maximize CAPWAP retries and timeout. Issues are detected less quickly, but the network is more resistant to minor transient packet drops.
- Enable SSH and configure credentials. Disable AP console.
- Enable AP monitor if needed but not radio monitor.
- Enable rogue detection and configure an RSSI threshold of -70 dBm.

Monitoring the Network

Once the network is up and running, you have to monitor it closely for problems. In a standard office environment users know the network and can either help each other in case of problems or open an internal helpdesk ticket. In a larger venue with many visitors come you want to focus on the biggest problems rather than specific individuals that can just have a misconfiguration, so you need to have the right monitoring strategy.

Monitoring the network from the Catalyst 9800 CLI or GUI is possible, but it is not the best tool to monitor daily. It is the most direct when you already have suspicions and/or data about the problem and want to run specific commands in real time. The main monitoring options are Cisco Catalyst Center or potentially a custom telemetry dashboard. It is possible to use 3rd party monitoring tools, but when those use SNMP as a protocol, the data is far from real-time and usual 3rd party monitoring tools are not granular enough with all the wireless vendors specificities. If you choose the SNMP protocol, make sure to use SNMPv3 as SNMPv2 has outdated security.

Cisco Catalyst Center is the best option as it allows you to manage your network on top of monitoring it. More than monitoring, it also allows to troubleshoot live and remediate many situations.

A custom telemetry dashboard can be useful if you want to display very specific metrics and widgets on a screen in an always-on fashion for a NOC or SOC. If there are very specific areas of your network that you want to keep an eye on, you could build dedicated widgets to show the network metrics in those areas in the manner of your choice.

For event networks, it is a good idea to monitor system wide RF statistics, in particular channel utilization

and number of clients per AP. This could be done from the CLI but provides only a snapshot at a specific point in time, channel utilization tends to be dynamic and is better suited for monitoring over time. For this type of monitoring a custom dashboard is usually a good approach. Other metrics that are more valuable when monitored over time can include, WNCD utilization, number of clients and their states, and venue specific metrics. An example of venue specific metrics would be monitoring usage and/or load for a specific area or location, for example hall *X* in the case of a conference center, or seating area *Y* in the case of an event venue.

For custom monitoring both NETCONF RPC (pull) and NETCONF streaming telemetry (push) are valid approaches, although using custom streaming telemetry in conjunction with Catalyst Center requires some diligence, as there is a limit to the number of telemetry subscriptions that can be configured on the WLC and Catalyst Center pre-populates (and utilizes) many of these.

When using NETCONF RPC some testing is required to ensure the WLC is not overloaded with NETCONF requests, particularly important to keep in mind are refresh rates for some of the data points and the time taken for the data to be returned. For example, AP channel utilization is refreshed (from AP to WLC) every 60 seconds, and collection of RF metrics for 1000 APs (from WLC) can take several seconds, in this example polling the WLC every 5 seconds would not be useful, a better approach would be to collect system-wide RF metrics every 3 minutes.

NETCONF is always the preferred over SNMP.

Lasty, monitoring of core network components cannot be overlooked, including DHCP pool utilization, number of NAT entries on core routers and so on. As the failure of any of these can easily be the cause of a *wireless* outage.

The harder you monitor, the more you can cause your own problems

There are a few classic scenarios where too much monitoring creates problems:

- Wireless sensors acting as clients can be nice to measure throughput and connectivity in specific places in the network, but remember that setting a high frequency of testing means the sensor will spend a lot of time doing large data transfers and therefore making the channel and network actually busy even if it was not otherwise. This is the prime example of "the harder you monitor, the worse your metrics look".
- SNMP, as already mentioned, is a legacy protocol that has a high CPU impact. Doing large SNMP polling can easily saturate your wireless network CPU with requests as soon as you frequently monitor all your wireless clients or APs when you have a large number of them. Always consider the amount of objects you are polling before setting an aggressive polling interval. The SNMP daemon is inside the IOSd process in IOS-XE, so when it is on high CPU it can impact the rest of IOS functionalities.
- Even if telemetry is more efficient than SNMP, large wireless networks can have a very high amount of clients, APs but also interferers and rogue devices. If your telemetry configuration is set too aggresively and the WLC has to report on any change in signal of any rogue device constantly, this can also easily saturate any controller appliance. Be sure to keep an eye on your "pubd" process that is in charge of telemetry and make sure it is not on high CPU utilization. If so, re-consider your thresholds and make sure to follow the 9800 best practices.

Issues Specific to Large Networks

If you have an SSID using web authentication, one problem can be clients that connect to that SSID and get an IP address but never authenticate because the end user is not actively trying to connect (the device connected automatically). The controller must intercept every HTTP packet sent by those clients that are in the state called *web authentication pending* and this uses WLC resources. Once your network is running, periodically keep an eye on the number of clients that are in web authentication pending state at a given time to see how it compares to baseline numbers. Same thing for clients in *IP Learn* state. You always have clients in that state when they are doing their DHCP process, but knowing what a proper working number for your network is helps to set a baseline and identify moments where this number can be too high and indicating a larger issue.

For large venues it is not uncommon to see ~10% of clients in *Web Auth Pending* state.

Day 2 Monitoring: Keeping an Eye on User Satisfaction

Once the network is up and running, there are two typical types of end user complains: they can't connect or struggle to connect (disconnections), or the Wi-Fi is operating slower than expected. The latter is very tricky to identify because it first depends on the expectations of the speed as well as the real-time density of a given area. Let's cover a few resources that can be helpful on your day-to-day monitoring of a large public venue network.

Validate Wi-Fi Throughput: Testing and Monitoring Guide. This cisco.com document covers how to monitor a network to spot for throughput issues. It goes through figuring out how much throughput clients can reasonably expect in your network when things are quiet and to estimate how much these estimates go down as client count and load increases. This is key to evaluate if an end user complaint about throughput is legitimate from a technical standpoint or not, and if you need to redesign that area for the load it faces potentially.

When clients report connectivity issues, after that was isolated and clarified with Catalyst Center, take a look at *Troubleshoot Catalyst 9800 Client Connectivity Issues Flow.*

Finally, as a general good practice, keep an eye on the overall key metrics of the WLC with the help of *Monitor Catalyst 9800 KPIs (Key Performance Indicators).*

Configuring for Scalability

SVIs and Interfaces on the 9800

Avoid creating SVIs for client VLANs on the WLC. Administrators used to older AireOS WLCs tend to have the reflex to create a layer 3 interface for each client VLAN, but this is rarely required. Interfaces increase the control plane attack vector and can require more ACLs with more complex entries. The WLC can be accessed, by default, on any of its interfaces, more work is needed to protect a WLC with more interfaces. It also complicates the routing, so it is best to avoid it.

Starting IOS XE 17.9, the SVI interfaces are no longer needed for mDNS snooping or DHCP relaying scenarios. There are therefore very few reasons to configure an SVI interface in a client VLAN.

Aggregated Probe Response

For large public networks, it is advisable to modify the default aggregate probe interval sent by access points. By default, the APs update the WLC every 500ms about the probes sent by clients. This information is used by load balancing, band select, location, and 802.11k features. If there are many clients and access points, it is advisable to modify the update interval to prevent control plane performance issues in the WLC. Recommended setting is 50 aggregated probe responses every 64 seconds. Also make sure that your APs are not reporting probes from locally administered MAC addresses as there is no point tracking those considering a single client could be using many locally administered MACs while scanning to avoid tracking on purpose.

```
wireless probe limit 50 64000

no wireless probe locally-administered-mac
```

IPv6

A lot of network administrators are still in denial of IPv6. There are only two acceptable options with IPv6: either you support it and must deploy adequate configuration everywhere, or you do not, and you must block it. It is not acceptable to not care about IPV6 and leave it enabled in some places without proper configuration. That would leave that whole IP world that your network security would be blind to.

If you enable IPv6, it is mandatory to configure a Virtual IPv6 address in the range 2001:DB8::/32 (that is an often-forgotten step).

It is important to note that, although IPv6 relies a lot on multicast for its basic operations, it can still operate if you disable multicast forwarding on the WLC. Multicast forwarding refers to client multicast data forwarding and not to the Neighbor Discovery, Router Solicitations, and other required protocols to operate IPv6.

If your internet connection or internet service provider provides IPv6 addresses, you can decide to allow IPv6 for your clients. That is a different decision from enabling IPv6 in your infrastructure. Your APs could keep operating in IPv4 only but still carry IPv6 client data traffic inside their CAPWAP packets. Enabling IPv6 on your infrastructure as well requires you to think about protecting client access to your APs, WLC and management subnet.

Verify the RA frequency of your client gateways. The WLC offers an RA throttling policy which limits the number of RAs forwarded to the clients as these can get chatty sometimes.

mDNS

In general, it is best to keep mDNS completely disabled in a large venue deployment.

mDNS bridging refers to the concept of allowing the mDNS packets to be sent as a Layer 2 multicast (therefore to the whole client subnet). mDNS became popular in home and small offices scenarios where it is very practical to discover services in your subnet. However, in a large network, this means sending the packet to all the clients in the subnet which is problematic from a traffic perspective in a large public network. On the other hand, bridging does not cause any overhead to the AP or WLC CPU as it is considered as regular data traffic. mDNS Proxy or mDNS gateway refers to the concept of using the WLC as a directory for all the services in the network. This allows to offer mDNS services across Layer 2 boundaries in an efficient manner and to also reduce overall traffic. With mDNS gateway, a printer, for example, sends its periodic service announcement via mDNS with a same-subnet Layer 2 multicast but the WLC does not forward it to all the other wireless clients. Instead, it takes a note of the service offered and registers it in its service directory. Whenever any client asks for services of a given type available, the WLC replies on behalf of the printer with the announcement. This avoids all the other wireless clients to hear about unnecessary requests and service offerings and only get a reply whenever they ask what services are around. While it greatly improves traffic efficiency, it does cause an overhead on the WLC (or the AP, if you rely on AP mDNS in FlexConnect scenarios) due to the snooping of mDNS traffic. If using mDNS gateway, it is critical to keep an eye on the CPU usage.

Bridging it leads to a multicast storm in your large subnet and snooping it (with the mDNS gateway feature) causes a lot of CPU utilization. Disable it globally as well as on each WLAN.

Some administrators enable mDNS because a couple of services need it in specific places, but it is important to understand how much unwanted traffic this adds. Apple devices are often advertising themselves as well as constantly hunting for services, causing a *background noise* of mDNS queries even when no one is making a particular use of any service. If you need to allow mDNS due to a certain business requirement, enable it globally and then enable it only on the WLAN where it is required and try to restrict the scope where mDNS is allowed.

Hardening the Network

Security

In large public networks, many things can be happening without the administrator knowing about it. People request cable drops in random places, or plug a home-grade switch in a location to have more switchports for their shenanigans, … They typically try these things without asking for permission first. This means that, even without a bad actor coming into play, security can already be compromised by good-willing customers and/or employees. It then becomes very easy for a bad actor to just walk around and find a cable to plug to and see what network access they get from there. Configuring 802.1X authentication on all switchports is a near-requirement for maintaining decent security in a large network. Catalyst Center can help you automate this rollout, and exceptions can be made for specific devices that do not support 802.1X authentication, but try to rely as little as possible on MAC-based authentication as that is (sincerely) not *real* security.

Rogue Access Points

Your strategy to combat rogues depends on a few factors. Many administrators instinctively go for very

strict rules, but the main questions are:

- When you get hundreds (if not thousands) of rogue alerts, do you have the human resources to look at all of them and take actions on all of them?
- Is your goal to physically remove the rogues to keep a clean RF spectrum? If so, you need many people to conduct this operation. Or maybe your goal is to only keep an eye on the security factor and just make sure the rogues do not represent any danger? This has a much more manageable human work cost.
- Enabling rogue detection can have an impact on your airtime and rogue containment has typically an even larger impact, did you analyze this impact and take it into account?

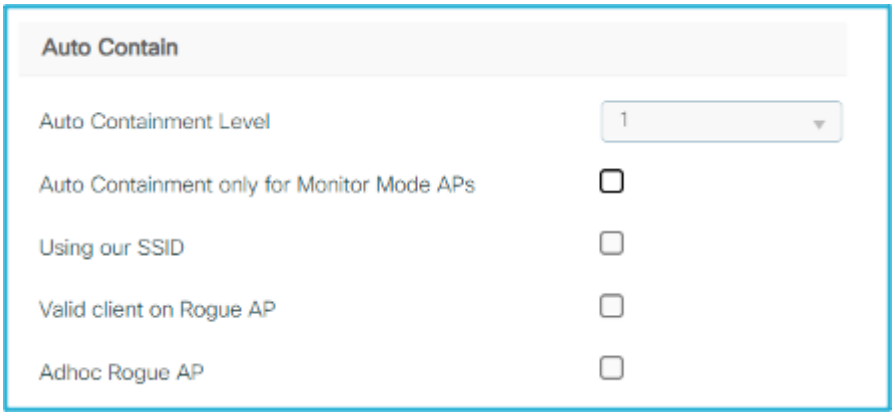
With regards to the impact of rogue detection, 9120 and 9130s have a dedicated CleanAir chip that takes care of the off-channel scanning (and therefore rogue detection) rendering the impact to the client-serving radio nearly null. 9160 series APs with their CleanAir Pro chip has a similar no-impact scanning capability but other APs that do not have the CleanAir chip need to take their client-serving radio off-channel to scan for rogues or to do containment. The AP model you are using therefore plays a role in the decision to use dedicated monitor-mode APs for rogue detection and containment or not.



Note: Mobile phones sharing a Wi-Fi hotspot operate in ‘infrastructure’ mode just like traditional APs, ‘ad-hoc” mode refers to a direct connection between mobile devices and is less common.

Rogue containment is often forbidden by regulatory rules, so it is essential that you check with your local authority before enabling it. Containing a rogue does not mean shutting down the rogue remotely but spamming the clients that try to connect to the rogue access point with deauthentication frames so that they do not connect. This can only work on legacy security SSID (it does not work in WPA3 or when PMF is enabled in WPA2) because your access points are not able to sign the deauthentication frames correctly. Containment has a negative impact on RF performance on the target channel as your APs are filling the airtime with deauthentication frames. It must therefore only be considered as a security measure to prevent your own legitimate clients to associate to a rogue access point by mistake. For all the reasons mentioned, it is recommended to not do any containment as it does not solve the rogue problem completely and causes more RF issues. If you need to use containment, it only makes sense to enable it for rogues that spoof one of your managed SSID as it is an obvious honeypot attack.

You can either configure auto containment with the “using our SSIDs” option:



You can also configure rogue rules to classify as malicious rogue access points according to your own criteria. Do not forget to enter the name of your neighboring and approved SSIDs as friendly rogues to remove those from your alarm list.

Enable AP authentication or PMF to protect your APs from impersonation.

A wired rogue is a rogue access point connected to your wired network, which is an increased security threat obviously. Detection of wired rogues is more complicated as the ethernet MAC address of a rogue typically differs from its radio MAC address. Cisco Catalyst Center has algorithms that still try to detect if a rogue is wired and hunts for rogue client MACs that are both heard over the air and seen on the wired infrastructure. The best solution to prevent wired rogues altogether is to secure all your switchports with 802.1X authentication.

If you are going to act physically on a rogue access point, leveraging Cisco Spaces is key to have an accurate location of the rogue. You most likely still need to search around once on site as people tend to hide rogue APs sometimes but reducing the search area to a few meters makes it a very feasible endeavor. Without Spaces, the rogue is shown on the map next to the AP detecting it the loudest which makes for a quite large search area. Many wireless tools and devices exist that show you the signal of the rogue access point in real time to help you locate the rogue physically.

Not exactly related to rogues, but since CleanAir was just covered, it is important to note that enabling CleanAir does not have a noticeable negative impact on performances except BLE beacon detection as this impacts 2.4GHz performance. You can configure your wireless to ignore Bluetooth interferers altogether as they are omnipresent in today’s world, and you cannot prevent your clients from enabling their Bluetooth.

There is another category of rogue that do not show up in your rogue detection list. Some clients can use Wi-Fi direct between each other and therefore do not use any access point. This does not really represent a security concern but does use RF air time.
One example is Apple AirDrop where two devices detect each other via Bluetooth and then can agree on using a Wi-Fi channel for file transfer. The users want to transfer this file anyway, so it is a good idea to let them transfer it as fast as possible so that the interference stops as soon as possible. However, be aware that airdrop detection can sometimes make a VPN tunnel flap if the VPN application doesn’t handle the AWDL interface status change correctly and to the end users it feels like it is your infrastructure Wi-Fi that is of poor quality when in reality it is the combination of their VPN application with the airdrop protocol that cause the flap. If disabling Bluetooth on the client solves the problem, it is not a problem on the Wi-Fi side.
Another well known example is Apple SideCar where you can use your iPad to extend your computer screen to another monitor. This can cause significant RF utilization as the streaming is continuous and can use a large channel width.

WiPS

WiPS covers more advanced attack vectors than just detecting the presence of a non-authorized rogue device. On top of those attacks, it also sometimes provides a PCAP of the event for Forensics analysis.

While this is a very useful security feature for the enterprise, a public-facing network must face the eternal question: what to do against it?

With the difficulty of managing many clients that you do not control, it is possible to divide the alarms into two categories. The alarms you can decide to ignore from Cisco Catalyst Center if you see too many of them are:

- 10001: DoS: Authentication Flood Alarm
- 10002: DoS: Association Request Alarm
- 10003: DoS: Broadcast Probe flood Alarm
- 10004: DoS: Disassociation Flood Alarm
- 10005: DoS: Broadcast Dis-Association Alarm
- 10006: DoS: De-authentication Flood Alarm
- 10007: DOS: Broadcast De-authentication Alarm
- 10008: DOS: EAPOL-Logoff Attack Alarm
- 10009: CTS flood alarm
- 10010: RTS Association Request Alarm
- 10011: Deauthentication Flood by Pair
- 10021: Airdrop Session (this one typically occurs a lot in any network and simply depicts regular peer to peer activity between Apple devices)
- 10022: Malformed Association Request
- 10023: Authentication Failure Flood by Signature
- 10024: Invalid MAC OUI by Signature
- 10025: Malformed Authentication

These alarms can potentially be caused by a misbehaving client. It is not possible to automatically prevent a denial-of-service attack since, essentially, you cannot prevent a faulty client from keeping the airtime busy. Even if the infrastructure ignores the client, it would still be able to use the medium and airtime to transmit, therefore impacting the performance of clients around it.

The other alarms are so specific that they most likely depict an actual malicious attack and can hardly happen due to bad client drivers. It is better to keep monitoring these alarms:

- 10012: Fuzzed beacon
- 10013: Fuzzed Probe request
- 10014: Fuzzed Probe response
- 10015: PS Poll Flood by Signature
- 10016: EAPOL Start V1 Flood by Signature
- 10017: Reassociation Request Flood by Destination
- 10018: Beacon Flood by Signature
- 10019: Probe Response Flood by Destination
- 10020: Block Ack Flood by Signature
- 10026/10027: RTS and CTS Virtual Carrier Sense Attack

The wireless infrastructure can sometimes take mitigation action like block listing the offending device, but the only real action to get rid of such an attack is to physically go there and remove the offending device.

It is advised to enable all forms of client exclusion to save airtime wasted by interacting with faulty clients.

Rate limiting clients

It is tempting for management and administrators to consider rate limiting users in order to prevent one user from hogging the bandwitdh in a specific location. However, it has been proven in the field many times that it is more efficient, from an aggregate air time standpoint, to not do any rate limiting. The faster the users get on the network and do their activities and the faster they are also off the network. Rate-limiting transfer speeds does not prevent a user to do large file transfers, it just makes it slower and drops packets in the process meaning the total air time wasted is even more typically than without any limitation. Rate limiting over wireless is done by dropping packets after the AP has forwarded a certain amount of packets in a certain time window coming from a client, if you do not do this per application you risk dropping important traffic from the client. But if you do this per application, it can have a different effect from what you want as certain applications use multiple connections and this can complicate throughput calculation for this application. Cisco CX experts advise to rather drop completely certain applications you do not want rather than use rate limiting. This is also covered in Cisco Wireless Meraki Cloud documentation : https://documentation.meraki.com/Wireless/Operate_and_Maintain/How_Tos/Firewall_and_Traffic_Shaping/Conf

as well as in Cisco Wireless Catalyst documentation https://www.cisco.com/c/en/us/support/docs/wireless/catalyst-9800-series-wireless-controllers/215441-configure-qos-rate-limiting-on-catalyst.html

Restricting Client Access

It is advised to enable peer-to-peer blocking on all your WLANs (unless you have a hard requirement for client-to-client communication - but this needs to be carefully considered and possibly limited). This feature prevents clients on the same WLAN from contacting each other. This is not a perfect solution as clients on different WLANs are able to still contact each other and clients belonging to different WLCs in the mobility group can also bypass this restriction. But it acts as an easy and efficient first layer of security and optimization. One more advantage of this feature of peer-to-peer blocking is that it also prevents client-to-client ARP which prevents applications from discovering other devices on the local network. Without peer-to-peer blocking, installing a simple application on the client could show all the other clients connected in the subnet with possibly their IP address and hostnames.

On top of this, it is recommended to apply both an IPv4 and an IPv6 (if you are using IPv6 in your network) ACL on your WLANs to prevent client-to-client communication. Applying an ACL that blocks client to client communication at the WLAN level works regardless of if you have client SVIs or not.

The other mandatory step is to preventing wireless client access to any form of management of your wireless controller.

Example:

```
ip access-list extended ACL_DENY_CLIENT_VLANS

 10 deny ip any 10.131.0.0 0.0.255.255

 20 deny ip 10.131.0.0 0.0.255.255 any
```

```
 30 deny ip any 10.132.0.0 0.0.255.255
 40 deny ip 10.132.0.0 0.0.255.255 any
 50 deny ip any 10.133.0.0 0.0.255.255
 60 deny ip 10.133.0.0 0.0.255.255 any
 70 deny ip any 10.134.0.0 0.0.255.255
 80 deny ip 10.134.0.0 0.0.255.255 any
 90 deny ip any 10.135.0.0 0.0.255.255
 100 deny ip 10.135.0.0 0.0.255.255 any
 110 deny ip any 10.136.0.0 0.0.255.255
 120 deny ip 10.136.0.0 0.0.255.255 any
 130 deny ip any 10.137.0.0 0.0.255.255
 140 deny ip 10.137.0.0 0.0.255.255 any
 150 permit ip any any
```

This ACL can be applied on the management interface SVI:

```
interface Vlan130
 ip access-group ACL_DENY_CLIENT_VLANS in
```

This is done on a WLC with client VLANS 131 to 137 created in the Layer 2 VLAN database but without any corresponding SVIs, and only one SVI exists for VLAN 130 which is how the WLC is managed. This ACL prevents all the wireless clients from sending any traffic to the WLC management and control planes completely. Don't forget that SSH or Web UI management is not the only thing you need to allow through, as a CAPWAP connection towards all APs is also required to be allowed. This is why this ACL has a default permit, but blocks wireless client ranges, rather than rely on a default deny all action which would require to specify all the allowed AP subnet ranges and management ranges.

Similarly, you can create another ACL that specifies all the possible management subnets:

```
ip access-list standard ACL_MGMT
 10 permit 10.128.0.0 0.0.255.255
 20 permit 10.127.0.0 0.0.255.255
 30 permit 10.100.0.0 0.0.255.255
 40 permit 10.121.0.0 0.0.255.255
```

```
 50 permit 10.141.0.0 0.0.255.255
```

You can then apply this ACL for CLI access:

```
line vty 0 50
 access-class ACL_MGMT in
 exec-timeout 180 0
 ipv6 access-class ACL_IPV6_MGMT in
 logging synchronous
 length 0
 transport preferred none
 transport input ssh
 transport output ssh
```

The same ACL can also be applied for web admin access.

Protecting from Traffic Storms

Multicasts and broadcasts are used more heavily by some applications than others. When considering a wired-only network, protecting against broadcast storm is often the only precaution taken. However, a multicast is as painful as a broadcast when sent over the air and it is important to understand why. First, imagine a packet sent (whether via broadcast or multicast) to all your wireless clients, that quickly adds up to many destinations. Each AP then needs to transmit this frame over the air in the most reliable way possible (although it is not guaranteed as reliable) and that is achieved by using a mandatory data rate (it is configurable which one of the mandatory rate is used). In layman's terms this means that frame is sent using an OFDM (802.11a/g) data rate, which is clearly not great.

In a large public network, it is not advised to rely on multicast to preserve airtime. However, in a large enterprise network you can have a requirement to keep multicast enabled for a specific application, although you must control it as much as possible to limit its impact. It is a good idea to document the application detail, multicast IP, and make sure to block other forms of multicast. Enabling Multicast forwarding is not a requirement for enabling IPv6, as explained previously. Broadcast forwarding is best kept disabled completely. Broadcasts are sometimes used by applications to discover other devices on the same subnet, which is clearly a security concern in a large network.

If you enable global multicast forwarding, make sure to use multicast-multicast AP CAPWAP setting. With this enabled, when the WLC receives a multicast packet from the wired infrastructure, it sends it to all interested APs with a single multicast packet, saving on a lot of packet duplication. Make sure to set a different CAPWAP multicast IP for each of your WLCs otherwise APs receive multicast traffic from other WLCs which is not desired.

If your APs are in other subnets from your Wireless Management Interface of the WLC (which is likely in a large network), you must enable multicast routing on your wired infrastructure. You can verify that all your APs are correctly receiving the multicast traffic with the command:

```
show ap multicast mom
```

IGMP (for IPv4 multicast) and MLD (for IPv6) multicast are also advised to be enabled in all cases if you need to rely on multicast. They allow only the interested wireless clients (and therefore only APs who have interested clients) to receive the multicast traffic. The WLC proxies the registration to the multicast traffic and takes care of keeping the registration alive, thereby offloading the clients.

Conclusion

Large public networks are complex, each one is unique with specific requirements and outcomes.

Respecting the guidelines in this document is a great starting point and helps to achieve success with your deployment while avoiding the most common issues. However, the guidelines are just guidelines and can require to be interpreted or adjusted within the context of the specific venue.

Cisco CX has teams of wireless professionals dedicated to large wireless deployments, with experience in numerous large events including sporting events and conferences. Reach out to your account team for further assistance.