

MODULE 1



Internet of Things

What Is IoT?

IoT is a technology transition in which devices will allow us to sense and control the physical world by making objects smarter and connecting them through an intelligent network.

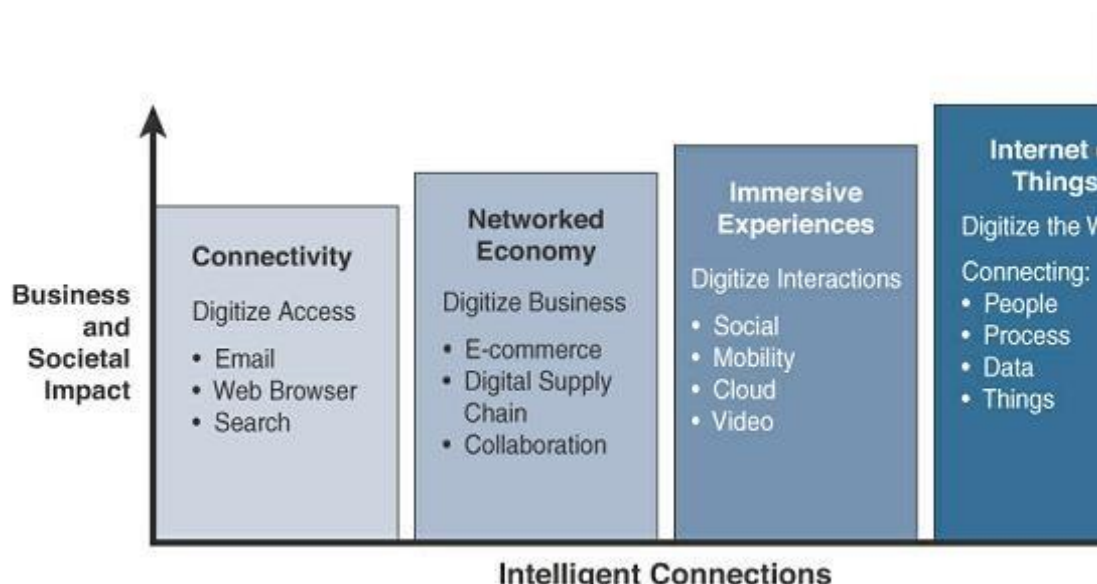
GOAL: The basic premise and goal of IoT is to “connect the unconnected.” This means that objects that are not currently joined to a computer network, namely the Internet, will be connected so that they can communicate and interact with people and other objects.

When objects and machines can be sensed and controlled remotely across a network, a tighter integration between the physical world and computers is enabled.

This allows for improvements in the areas of efficiency, accuracy, automation, and the enablement of advanced applications.

GENESIS OF IOT

The person credited with the creation of the term “Internet of Things” is Kevin Ashton. While working for Procter & Gamble in 1999, Kevin used this phrase to explain a new idea related to linking the company’s supply chain to the Internet.



the evolution of the Internet can be categorized into four phases. Each of these phases has had a profound impact on our society and our lives. These four phases are further defined in Table below.

Internet of Things

Internet Phase	Definition
Connectivity (Digitize access)	This phase connected people to email, web services, and search so that information is easily accessed.
Networked Economy (Digitize business)	This phase enabled e-commerce and supply chain enhancements along with collaborative engagement to drive increased efficiency in business processes.
Immersive Experiences (Digitize interactions)	This phase extended the Internet experience to encompass widespread video and social media while always being connected through mobility. More and more applications are moved into the cloud.
Internet of Things (Digitize the world)	This phase is adding connectivity to objects and machines in the world around us to enable new services and experiences. It is connecting the unconnected.

IOT AND DIGITIZATION

IoT and *digitization* are terms that are often used interchangeably. In most contexts, this duality is fine, but there are key differences to be aware of.

At a high level, IoT focuses on connecting “things,” such as objects and machines, to a computer network, such as the Internet. IoT is a well-understood term used across the industry as a whole. On the other hand, digitization can mean different things to different people but generally encompasses the connection of “things” with the data they generate and the business insights that result.

Digitization, as defined in its simplest form, is the conversion of information into a digital format. Digitization has been happening in one form or another for several decades. For example, the whole photography industry has been digitized. Pretty much everyone has digital cameras these days, either standalone devices or built into their mobile phones. Almost no one buys film and takes it to a retailer to get it developed. The digitization of photography has completely changed our experience when it comes to capturing images.

CONVERGENCE OF IT AND OT

Until recently, information technology (IT) and operational technology (OT) have for the most part lived in separate worlds. IT supports connections to the Internet along with related data and technology systems and is focused on the secure flow of data across an organization. OT monitors and controls devices and processes on physical operational systems. These systems include assembly lines, utility distribution networks, production facilities, roadway systems, and many more. Typically, IT did not get involved with the production and logistics of OT environments.

Management of OT is tied to the lifeblood of a company. For example, if the network connecting the machines in a factory fails, the machines cannot function, and production may come to a standstill, negatively impacting business on the order of millions of dollars. On the other hand, if the email server (run by the IT department) fails for a few hours, it may irritate people, but it is unlikely to impact business at anywhere near the same level. **Table below highlights some of the differences between IT and OT networks and their various challenges.**

Internet of Things

Criterion	Industrial OT Network	Enterprise IT Network
Operational focus	Keep the business operating 24x7	Manage the computers, data, and employee communication system in a secure way
Priorities	1. Availability 2. Integrity 3. Security	1. Security 2. Integrity 3. Availability
Types of data	Monitoring, control, and supervisory data	Voice, video, transactional, and bulk data
Security	Controlled physical access to devices	Devices and users authenticated to the network
Implication of failure	OT network disruption directly impacts business	Can be business impacting, depending on industry, but workarounds may be possible
Network upgrades (software or hardware)	Only during operational maintenance windows	Often requires an outage window when workers are not onsite; impact can be mitigated
Security vulnerability	Low: OT networks are isolated and often use proprietary protocols	High: continual patching of hosts is required, and the network is connected to Internet and requires vigilant protection

Source: Maciej Kranz, *IT Is from Venus, OT Is from Mars*, blogs.cisco.com/digital/it-is-from-venus-ot-is-from-mars, July 14, 2015.

IOT CHALLENGES

The most significant challenges and problems that IoT is currently facing are

Challenge	Description
Scale	While the scale of IT networks can be large, the scale of OT can be several orders of magnitude larger. For example, one large electrical utility in Asia recently began deploying IPv6-based smart meters on its electrical grid. While this utility company has tens of thousands of employees (which can be considered IP nodes in the network), the number of meters in the service area is tens of millions. This means the scale of the network the utility is managing has increased by more than 1,000-fold! Chapter 5, “IP as the IoT Network Layer,” explores how new design approaches are being developed to scale IPv6 networks into the millions of devices.
Security	With more “things” becoming connected with other “things” and people, security is an increasingly complex issue for IoT. Your threat surface is now greatly expanded, and if a device gets hacked, its connectivity is a major concern. A compromised device can serve as a launching point to attack other devices and systems. IoT security is also pervasive across just about every facet of IoT. For more information on IoT security, see Chapter 8, “Securing IoT.”

Internet of Things

Privacy	As sensors become more prolific in our everyday lives, much of the data they gather will be specific to individuals and their activities. This data can range from health information to shopping patterns and transactions at a retail establishment. For businesses, this data has monetary value. Organizations are now discussing who owns this data and how individuals can control whether it is shared and with whom.
Big data and data analytics	IoT and its large number of sensors is going to trigger a deluge of data that must be handled. This data will provide critical information and insights if it can be processed in an efficient manner. The challenge, however, is evaluating massive amounts of data arriving from different sources in various forms and doing so in a timely manner.
Interoperability	As with any other nascent technology, various protocols and architectures are jockeying for market share and standardization within IoT. Some of these protocols and architectures are based on proprietary elements, and others are open. Recent IoT standards are helping minimize this problem, but there are often various protocols and implementations available for IoT networks. The prominent protocols and architectures—especially open, standards-based implementations—are the subject of this book.

IoT Network Architecture and Design

The unique challenges posed by IoT networks and how these challenges have driven new architectural models.

- Drivers Behind New Network Architectures
- Comparing IoT Architectures.
- A Simplified IoT Architecture
- The Core IoT Functional Stack
- IoT Data Management and Compute Stack

DRIVERS BEHIND NEW NETWORK ARCHITECTURES

This begins by comparing how using an architectural blueprint to construct a house is similar to the approach we take when designing a network. Take a closer look at some of the differences between IT and IoT networks, with a focus on the IoT requirements that are driving new network architectures, and considers what adjustments are needed.

Internet of Things

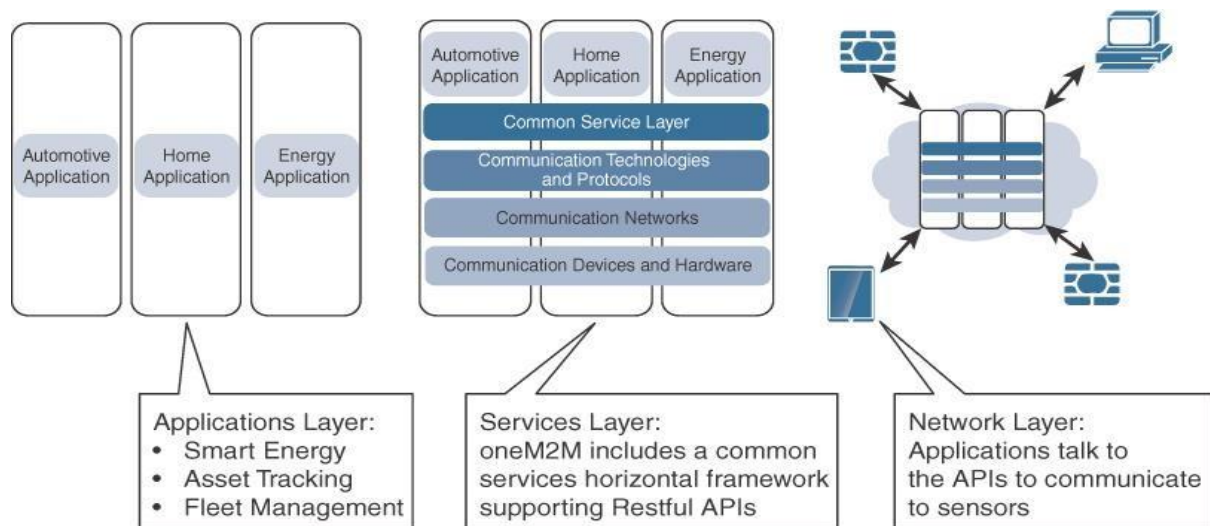
Challenge	Description	IoT Architectural Change Required
Scale	The massive scale of IoT endpoints (sensors) is far beyond that of typical IT networks.	The IPv4 address space has reached exhaustion and is unable to meet IoT's scalability requirements. Scale can be met only by using IPv6. IT networks continue to use IPv4 through features like Network Address Translation (NAT).
Security	IoT devices, especially those on wireless sensor networks (WSNs), are often physically exposed to the world.	Security is required at every level of the IoT network. Every IoT endpoint node on the network must be part of the overall security strategy and must support device-level authentication and link encryption. It must also be easy to deploy with some type of a zero-touch deployment model.
Devices and networks constrained by power, CPU, memory, and link speed	Due to the massive scale and longer distances, the networks are often constrained, lossy, and capable of supporting only minimal data rates (tens of bps to hundreds of Kbps).	New last-mile wireless technologies are needed to support constrained IoT devices over long distances. The network is also constrained, meaning modifications need to be made to traditional network-layer transport mechanisms.
The massive volume of data generated	The sensors generate a massive amount of data on a daily basis, causing network bottlenecks and slow analytics in the cloud.	Data analytics capabilities need to be distributed throughout the IoT network, from the edge to the cloud. In traditional IT networks, analytics and applications typically run only in the cloud.
Support for legacy devices	An IoT network often comprises a collection of modern, IP-capable endpoints as well as legacy, non-IP devices that rely on serial or proprietary protocols.	Digital transformation is a long process that may take many years, and IoT networks need to support protocol translation and/or tunneling mechanisms to support legacy protocols over standards-based protocols, such as Ethernet and IP.
The need for data to be analyzed in real time	Whereas traditional IT networks perform scheduled batch processing of data, IoT data needs to be analyzed and responded to in real-time.	Analytics software needs to be positioned closer to the edge and should support real-time streaming analytics. Traditional IT analytics software (such as relational databases or even Hadoop), are better suited to batch-level analytics that occur after the fact.

COMPARING IOT ARCHITECTURES

The oneM2M IoT Standardized Architecture

In an effort to standardize the rapidly growing field of machine-to-machine (M2M) communications, the European Telecommunications Standards Institute (ETSI) created the M2M Technical Committee in 2008. The goal of this committee was to create a common architecture that would help accelerate the adoption of M2M applications and devices. Over time, the scope has expanded to include the Internet of Things.

One of the greatest challenges in designing an IoT architecture is dealing with the heterogeneity of devices, software, and access methods. By developing a horizontal platform architecture, oneM2M is developing standards that allow interoperability at all levels of the IoT stack



The Main Elements of the oneM2M IoT Architecture

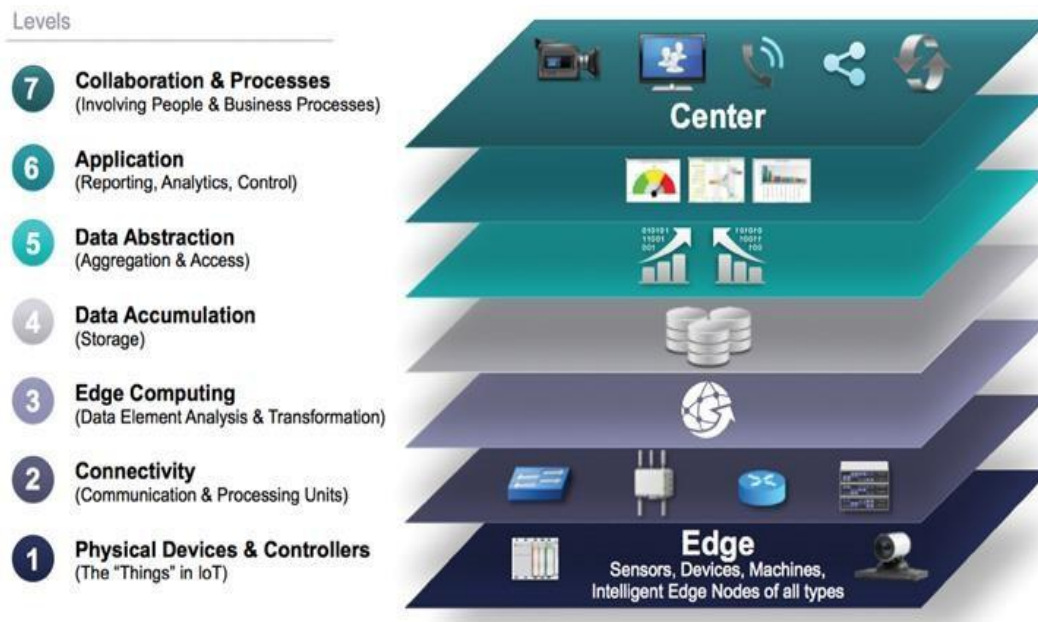
The oneM2M architecture divides IoT functions into three major domains: the application layer, the services layer, and the network layer

- **Applications layer:** The oneM2M architecture gives major attention to connectivity between devices and their applications. This domain includes the application-layer protocols and attempts to standardize northbound API definitions for interaction with business intelligence (BI) systems. Applications tend to be industry-specific and have their own sets of data models, and thus they are shown as vertical entities.
- **Services layer:** This layer is shown as a horizontal framework across the vertical industry applications. At this layer, horizontal modules include the physical network that the IoT applications run on, the underlying management protocols, and the hardware. Examples include backhaul communications via cellular, MPLS networks, VPNs, and so on. Riding on top is the common services layer.
- **Network layer:** This is the communication domain for the IoT devices and endpoints. It includes the devices themselves and the communications network that links them. Embodiments of this communications infrastructure include wireless mesh technologies, such as IEEE 802.15.4, and wireless point-to-multipoint systems, such as IEEE 801.11ah.

The IoT World Forum (IoTWF) Standardized Architecture

This publish a seven-layer IoT architectural reference model.

- While various IoT reference models exist, the one put forth by the IoT World Forum offers a clean, simplified perspective on IoT and includes edge computing, data storage, and access. It provides a succinct way of visualizing IoT from a technical perspective. Each of the seven layers is broken down into specific functions, and security encompasses the entire model.



Using this reference model, we are able to achieve the following:

1. Decompose the IoT problem into smaller parts
2. Identify different technologies at each layer and how they relate to one another
3. Define a system in which different parts can be provided by different vendors
4. Have a process of defining interfaces that leads to interoperability
5. Define a tiered security model that is enforced at the transition points between levels

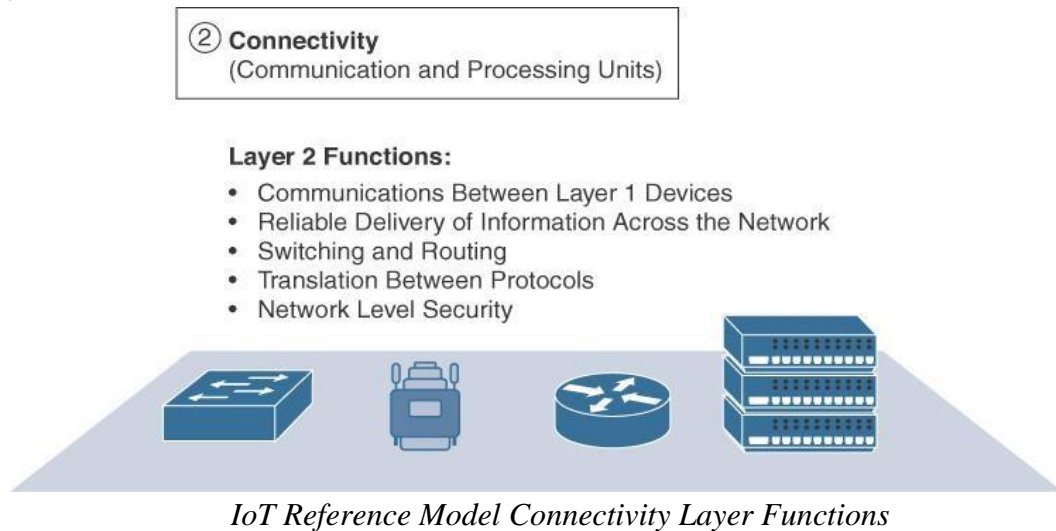
Layer 1: Physical Devices and Controllers Layer

The first layer of the IoT Reference Model is the physical devices and controllers layer. This layer is home to the “things” in the Internet of Things, including the various endpoint devices and sensors that send and receive information. The size of these “things” can range from almost microscopic sensors to giant machines in a factory. Their primary function is generating data and being capable of being queried and/or controlled over a network.

Internet of Things

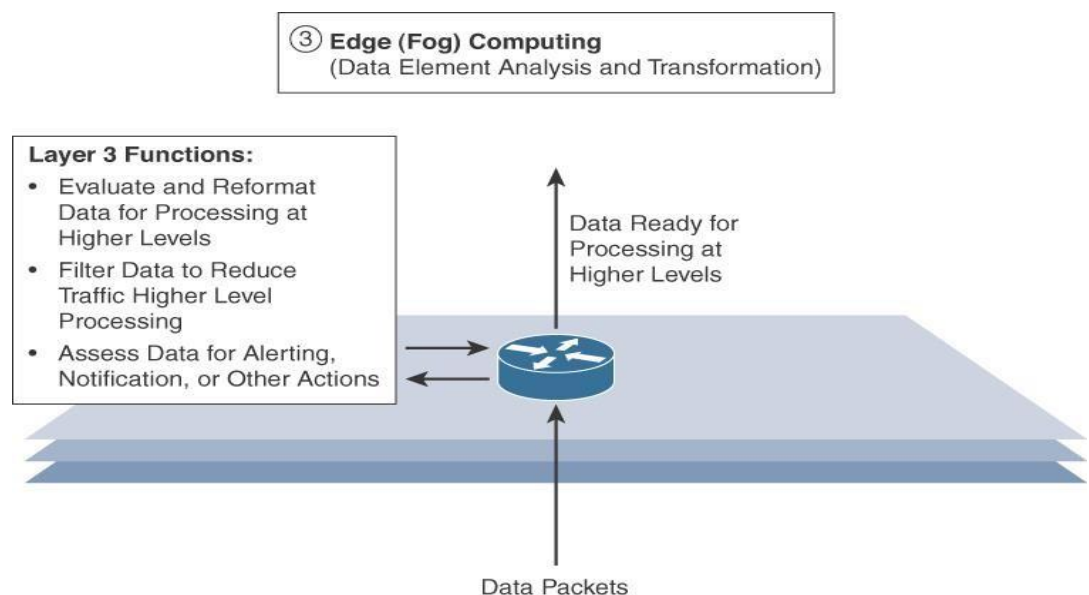
Layer 2: Connectivity Layer

In the second layer of the IoT Reference Model, the focus is on connectivity. The most important function of this IoT layer is the reliable and timely transmission of data. More specifically, this includes transmissions between Layer 1 devices and the network and between the network and information processing that occurs at Layer 3 (the edge computing layer).



Layer 3: Edge Computing Layer

Edge computing is the role of Layer 3. Edge computing is often referred to as the “fog” layer and is discussed in the section “[Fog Computing](#),” later in this chapter. At this layer, the emphasis is on data reduction and converting network data flows into information that is ready for storage and processing by higher layers. One of the basic principles of this reference model is that information processing is initiated as early and as close to the edge of the network as possible



Internet of Things

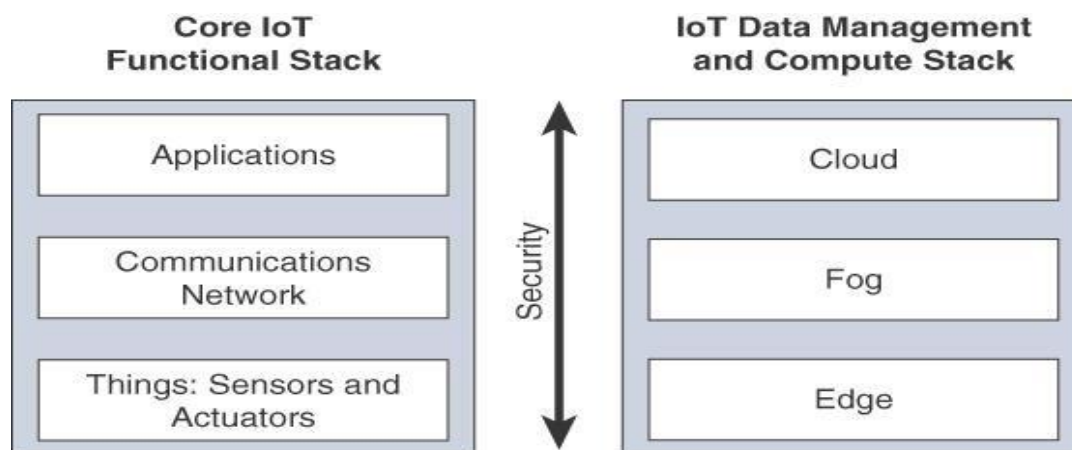
Another important function that occurs at Layer 3 is the evaluation of data to see if it can be filtered or aggregated before being sent to a higher layer. This also allows for data to be reformatted or decoded, making additional processing by other systems easier. Thus, a critical function is assessing the data to see if predefined thresholds are crossed and any action or alerts need to be sent.

Upper Layers: Layers 4–7

The upper layers deal with handling and processing the IoT data generated by the bottom layer. For the sake of completeness, Layers 4–7 of the IoT Reference Model are summarized in [Table .](#)

IoT Reference Model Layer	Functions
Layer 4: Data accumulation layer	Captures data and stores it so it is usable by applications when necessary. Converts event-based data to query-based processing.
Layer 5: Data abstraction layer	Reconciles multiple data formats and ensures consistent semantics from various sources. Confirms that the data set is complete and consolidates data into one place or multiple data stores using virtualization.
Layer 6: Applications layer	Interprets data using software applications. Applications may monitor, control, and provide reports based on the analysis of the data.
Layer 7: Collaboration and processes layer	Consumes and shares the application information. Collaborating on and communicating IoT information often requires multiple steps, and it is what makes IoT useful. This layer can change business processes and delivers the benefits of IoT.

A SIMPLIFIED IOT ARCHITECTURE

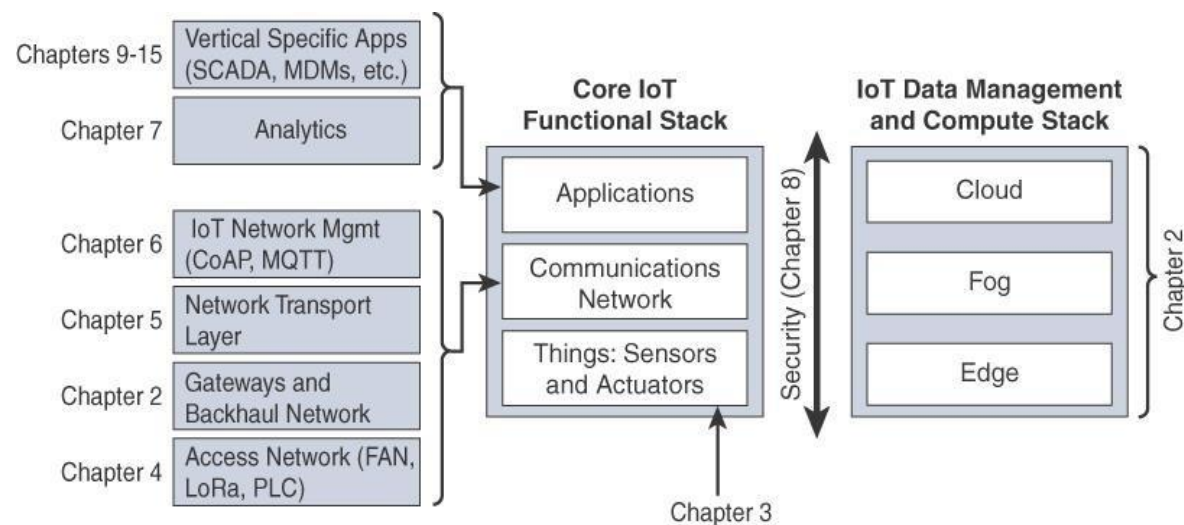


Simplified IoT Architecture

Internet of Things

The presentation of the Core IoT Functional Stack in three layers is meant to simplify your understanding of the IoT architecture into its most foundational building blocks. The network communications layer of the IoT stack itself involves a significant amount of detail and incorporates a vast array of technologies.

Data management is aligned with each of the three layers of the Core IoT Functional Stack. The three data management layers are the edge layer (data management within the sensors themselves), the fog layer (data management in the gateways and transit network), and the cloud layer (data management in the cloud or central data center). An expanded view of the IoT architecture presented as below:



The Core IoT Functional Stack can be expanded into sublayers containing greater detail and specific network functions. For example, the communications layer is broken down into four separate sublayers: the access network, gateways and backhaul, IP transport, and operations and management sublayers.

The applications layer of IoT networks is quite different from the application layer of a typical enterprise network. Instead of simply using business applications, IoT often involves a strong big data analytics component. One message that is stressed throughout this book is that IoT is not just about the control of IoT devices but, rather, the useful insights gained from the data generated by those devices. Thus, the applications layer typically has both analytics and industry-specific IoT control system components.

presented in [Part II](#), and it gives you the tools you need to understand how these technologies are applied in key industries in [Part III](#).

THE CORE IOT FUNCTIONAL STACK

IoT networks are built around the concept of “things,” or smart objects performing functions and delivering new connected services. These objects are “smart” because they use a combination of contextual information and configured goals to perform actions.

From an architectural standpoint, several components have to work together for an IoT network to be operational:

Internet of Things

- “Things” layer:
- Communications network layer
- Access network sublayer
- Gateways and backhaul network sublayer
- Network transport sublayer
- IoT network management sublayer
- Application and analytics layer

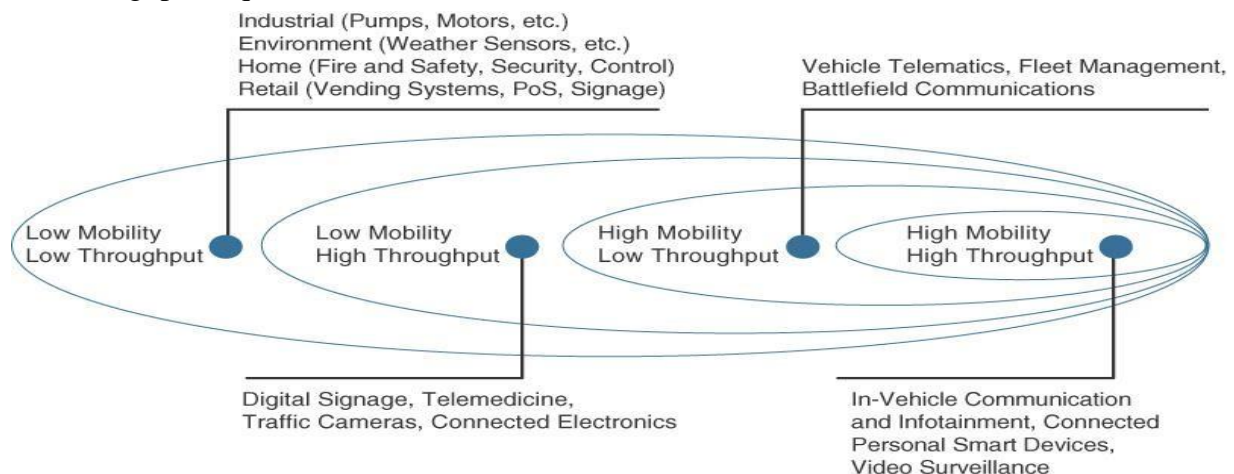
The following sections examine these elements and help you architect your IoT communication network.

Layer 1: Things: Sensors and Actuators Layer

“Smart Objects: The ‘Things’ in IoT,” provides more in-depth information about smart objects. From an architectural standpoint, the variety of smart object types, shapes, and needs drive the variety of IoT protocols and architectures. One architectural classification could be:

- **Battery-powered or power-connected:** This classification is based on whether the object carries its own energy supply or receives continuous power from an external power source.
- **Mobile or static:** This classification is based on whether the “thing” should move or always stay at the same location. A sensor may be mobile because it is moved from one object to another or because it is attached to a moving object.
- **Low or high reporting frequency:** This classification is based on how often the object should report monitored parameters. A rust sensor may report values once a month. A motion sensor may report acceleration several hundred times per second.
- **Simple or rich data:** This classification is based on the quantity of data exchanged at each report cycle.
- **Report range:** This classification is based on the distance at which the gateway is located. For example, for your fitness band to communicate with your phone, it needs to be located a few meters away at most.
- **Object density per cell:** This classification is based on the number of smart objects (with a similar need to communicate) over a given area, connected to the same gateway.

Below figure provides some examples of applications matching the combination of mobility and throughput requirements.



Example of Sensor Applications Based on Mobility and Throughput

Internet of Things

Layer 2: Communications Network Layer

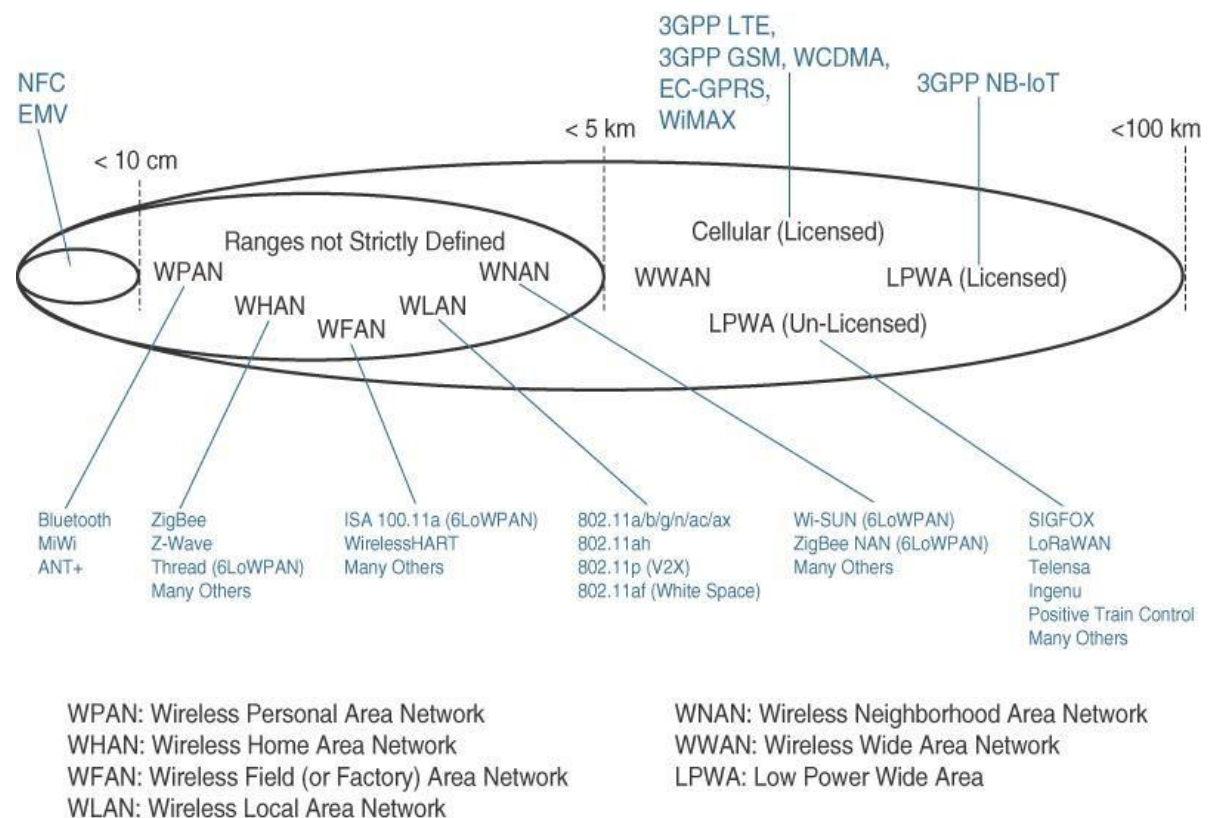
Once you have determined the influence of the smart object form factor over its transmission capabilities (transmission range, data volume and frequency, sensor density and mobility), you are ready to connect the object and communicate.

Compute and network assets used in IoT can be very different from those in IT environments. The difference in the physical form factors between devices used by IT and OT is obvious even to the most casual of observers. What typically drives this is the physical environment in which the devices are deployed. What may not be as inherently obvious, however, is their operational differences. The operational differences must be understood in order to apply the correct handling to secure the target assets.

Access Network Sublayer

There is a direct relationship between the IoT network technology you choose and the type of connectivity topology this technology allows. Each technology was designed with a certain number of use cases in mind (what to connect, where to connect, how much data to transport at what interval and over what distance). These use cases determined the frequency band that was expected to be most suitable, the frame structure matching the expected data pattern (packet size and communication intervals), and the possible topologies that these use cases illustrate.

One key parameter determining the choice of access technology is the range between the smart object and the information collector. [Figure 2-9](#) lists some access technologies you may encounter in the IoT world and the expected transmission distances.



Internet of Things

- ✓ Range estimates are grouped by category names that illustrate the environment or the vertical where data collection over that range is expected. Common groups are as follows:

■ **PAN (personal area network):** Scale of a few meters. This is the personal space around a person. A common wireless technology for this scale is Bluetooth.

■ **HAN (home area network):** Scale of a few tens of meters. At this scale, common wireless technologies for IoT include ZigBee and Bluetooth Low Energy (BLE).

■ **NAN (neighborhood area network):** Scale of a few hundreds of meters. The term NAN is often used to refer to a group of house units from which data is collected.

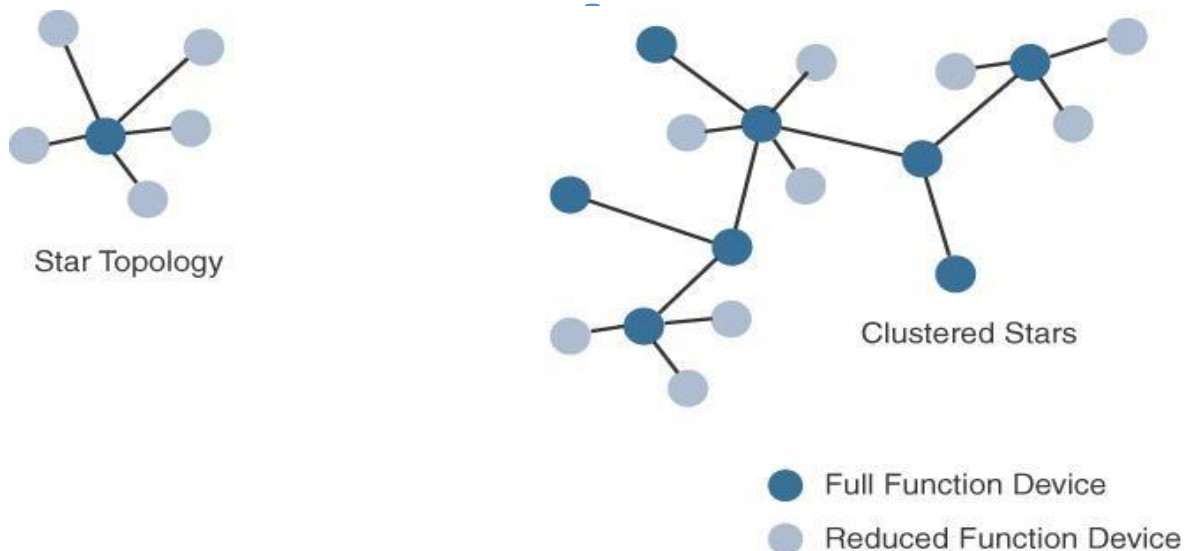
■ **FAN (field area network):** Scale of several tens of meters to several hundred meters. FAN typically refers to an outdoor area larger than a single group of house units. The FAN is often seen as “open space” (and therefore not secured and not controlled).

■ **LAN (local area network):** Scale of up to 100 m. This term is very common in networking, and it is therefore also commonly used in the IoT space when standard networking technologies (such as Ethernet or IEEE 802.11) are used.

- ✓ Similar ranges also do not mean similar topologies. Some technologies offer flexible connectivity structure to extend communication possibilities:

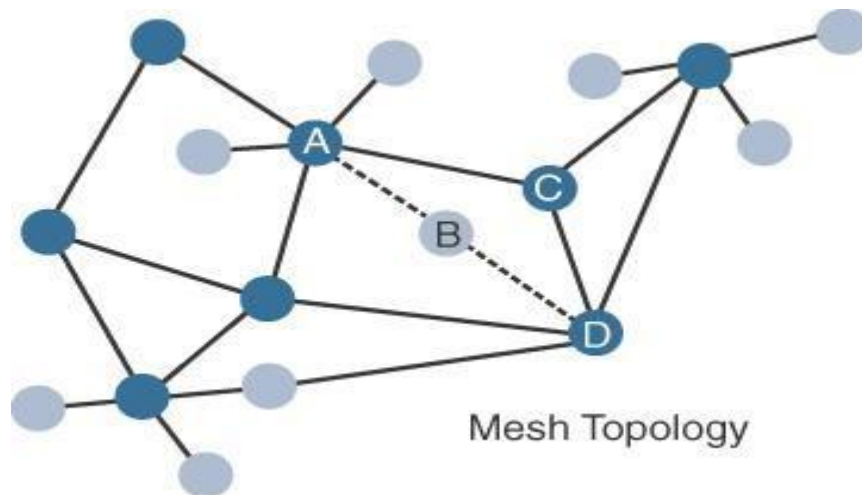
■ **Point-to-point topologies**

■ **Point-to-multipoint**



Star and Clustered Star Topologies

Internet of Things



Comparison of the main solutions from an architectural angle.

Technology	Type and Range	Architectural Characteristics
Ethernet	Wired, 100 m max	Requires a cable per sensor/sensor group; adapted to static sensor position in a stable environment; range is limited; link is very reliable
Wi-Fi (2.4 GHz, 5 GHz)	Wireless, 100 m (multipoint) to a few kilometers (P2P)	Can connect multiple clients (typically fewer than 200) to a single AP; range is limited; adapted to cases where client power is not an issue (continuous power or client battery recharged easily); large bandwidth available, but interference from other systems likely; AP needs a cable
802.11ah (HaloW, Wi-Fi in sub-1 GHz)	Wireless, 1.5 km (multipoint), 10 km (P2P)	Can connect a large number of clients (up to 6000 per AP); longer range than traditional Wi-Fi; power efficient; limited bandwidth; low adoption; and cost may be an issue
WiMAX (802.16)	Wireless, several kilometers (last mile), up to 50 km (backhaul)	Can connect a large number of clients; large bandwidth available in licensed spectrum (fee-based); reduced bandwidth in license-free spectrum (interferences from other systems likely); adoption varies on location
Cellular (for example, LTE)	Wireless, several kilometers	Can connect a large number of clients; large bandwidth available; licensed spectrum (interference-free; license-based)

Architectural Considerations for WiMAX and Cellular Technologies

Layer 3: Applications and Analytics Layer

Once connected to a network, your smart objects exchange information with other systems. As soon as your IoT network spans more than a few sensors, the power of the Internet of Things appears in the applications that make use of the information exchanged with the smart objects.

Internet of Things

Analytics Versus Control Applications

Multiple applications can help increase the efficiency of an IoT network. Each application collects data and provides a range of functions based on analyzing the collected data. It can be difficult to compare the features offered. From an architectural standpoint, one basic classification can be as follows:

■ **Analytics application:** This type of application collects data from multiple smart objects, processes the collected data, and displays information resulting from the data that was processed. The display can be about any aspect of the IoT network, from historical reports, statistics, or trends to individual system states. The important aspect is that the application processes the data to convey a view of the network that cannot be obtained from solely looking at the information displayed by a single smart object.

■ **Control application:** This type of application controls the behavior of the smart object or the behavior of an object related to the smart object. For example, a pressure sensor may be connected to a pump. A control application increases the pump speed when the connected sensor detects a drop in pressure. Control applications are very useful for controlling complex aspects of an IoT network with a logic that cannot be programmed inside a single IoT object, either because the configured changes are too complex to fit into the local system or because the configured changes rely on parameters that include elements outside the IoT object.

Data Versus Network Analytics

Analytics is a general term that describes processing information to make sense of collected data. In the world of IoT, a possible classification of the analytics function is as follows:

■ **Data analytics:** This type of analytics processes the data collected by smart objects and combines it to provide an intelligent view related to the IoT system. At a very basic level, a dashboard can display an alarm when a weight sensor detects that a shelf is empty in a store. In a more complex case, temperature, pressure, wind, humidity, and light levels collected from thousands of sensors may be combined and then processed to determine the likelihood of a storm and its possible path .

■ **Network analytics:** Most IoT systems are built around smart objects connected to the network. A loss or degradation in connectivity is likely to affect the efficiency of the system. Such a loss can have dramatic effects. For example, open mines use wireless networks to automatically pilot dump trucks. A lasting loss of connectivity may result in an accident or degradation of operations efficiency (automated dump trucks typically stop upon connectivity loss). On a more minor scale, loss of connectivity means that data stops being fed to your data analytics platform, and the system stops making intelligent analyses of the IoT system.

Data Analytics Versus Business Benefits

Data analytics is undoubtedly a field where the value of IoT is booming. Almost any object can be connected, and multiple types of sensors can be installed on a given object. Collecting and interpreting the data generated by these devices is where the value of IoT is realized.

Smart Services

- The ability to use IoT to improve operations is often termed “smart services.” This term is generic, and in many cases the term is used but its meaning is often stretched to include one form of service or another where an additional level of intelligence is provided.

Internet of Things

- Smart services can also be used to measure the efficiency of machines by detecting machine output, speed, or other forms of usage evaluation.
- Smart services can be integrated into an IoT system. For example, sensors can be integrated in a light bulb. A sensor can turn a light on or off based on the presence of a human in the room.

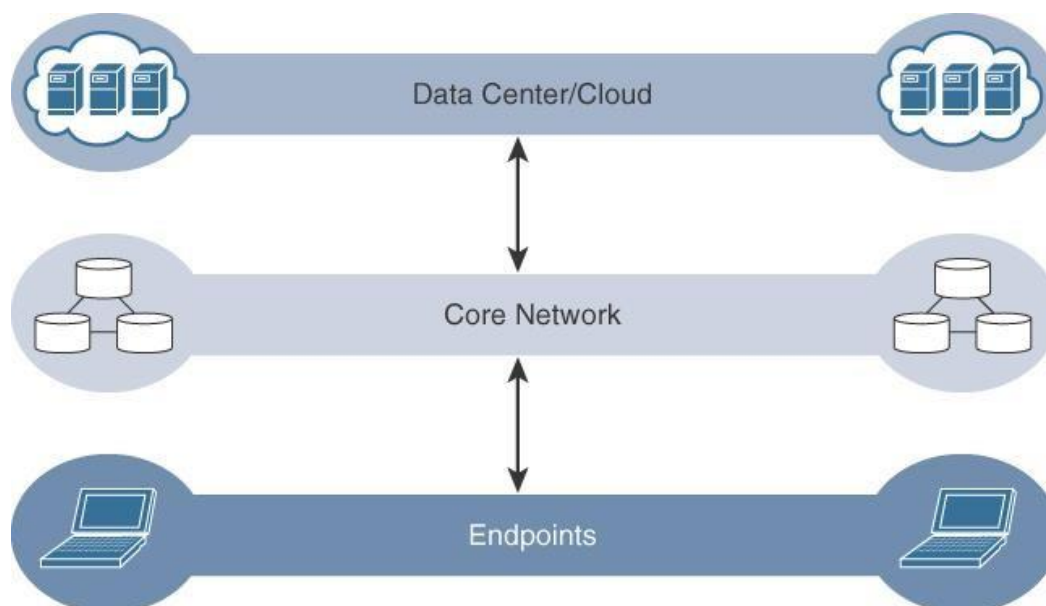
IOT DATA MANAGEMENT AND COMPUTE STACK

This model also has limitations. As data volume, the variety of objects connecting to the network, and the need for more efficiency increase, new requirements appear, and those requirements tend to bring the need for data analysis closer to the IoT system. These new requirements include the following:

■ **Minimizing latency:** Milliseconds matter for many types of industrial systems, such as when you are trying to prevent manufacturing line shutdowns or restore electrical service. Analyzing data close to the device that collected the data can make a difference between averting disaster and a cascading system failure.

■ **Conserving network bandwidth:** Offshore oil rigs generate 500 GB of data weekly. Commercial jets generate 10 TB for every 30 minutes of flight. It is not practical to transport vast amounts of data from thousands or hundreds of thousands of edge devices to the cloud. Nor is it necessary because many critical analyses do not require cloud-scale processing and storage.

■ **Increasing local efficiency:** Collecting and securing data across a wide geographic area with different environmental conditions may not be useful. The environmental conditions in one area will trigger a local response independent from the conditions of another site hundreds of miles away. Analyzing both areas in the same cloud system may not be necessary for immediate efficiency.



The Traditional IT Cloud Computing Model

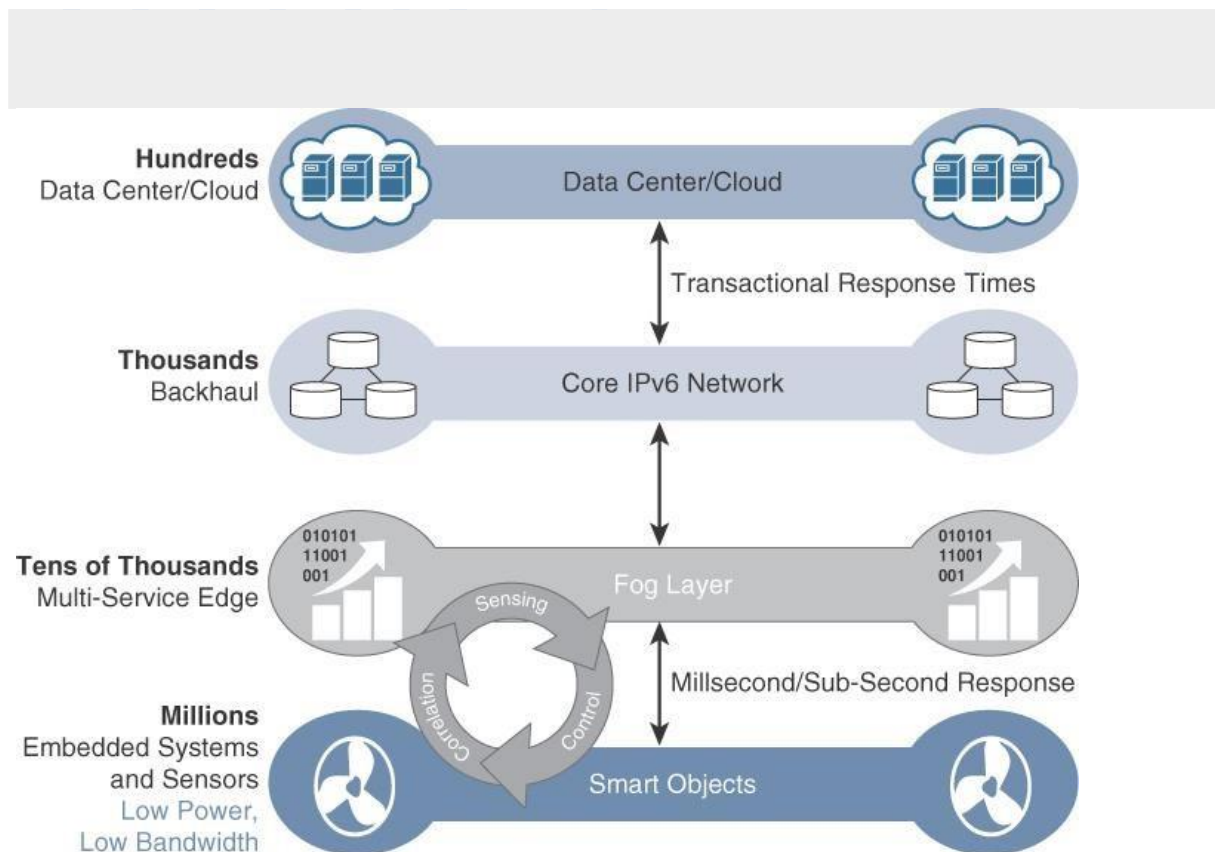
Internet of Things

IoT systems function differently. Several data-related problems need to be addressed:

- Bandwidth in last-mile IoT networks is very limited. When dealing with thousands/millions of devices, available bandwidth may be on order of tens of Kbps per device or even less.
- Latency can be very high. Instead of dealing with latency in the milliseconds range, large IoT networks often introduce latency of hundreds to thousands of milliseconds.
- Network backhaul from the gateway can be unreliable and often depends on 3G/LTE or even satellite links. Backhaul links can also be expensive if a per-byte data usage model is necessary.
- The volume of data transmitted over the backhaul can be high, and much of the data may not really be that interesting (such as simple polling messages).
- Big data is getting bigger. The concept of storing and analyzing all sensor data in the cloud is impractical. The sheer volume of data generated makes real-time analysis and response to the data almost impossible.

Fog Computing

The solution to the challenges mentioned in the previous section is to distribute data management throughout the IoT system, as close to the edge of the IP network as possible. The best-known embodiment of edge services in IoT is fog computing. Any device with computing, storage, and network connectivity can be a fog node. Examples include industrial controllers, switches, routers, embedded servers, and IoT gateways. Analyzing IoT data close to where it is collected minimizes latency, offloads gigabytes of network traffic from the core network, and keeps sensitive data inside the local network.



The IoT Data Management and Compute Stack with Fog Computing

Internet of Things

Fog services are typically accomplished very close to the edge device, sitting as close to the IoT endpoints as possible. One significant advantage of this is that the fog node has contextual awareness of the sensors it is managing because of its geographic proximity to those sensors. For example, there might be a fog router on an oil derrick that is monitoring all the sensor activity at that location. Because the fog node is able to analyze information from all the sensors on that derrick, it can provide contextual analysis of the messages it is receiving and may decide to send back only the relevant information over the backhaul network to the cloud. In this way, it is performing distributed analytics such that the volume of data sent upstream is greatly reduced and is much more useful to application and analytics servers residing in the cloud.

Fog applications are as diverse as the Internet of Things itself. What they have in common is data reduction—monitoring or analyzing real-time data from network-connected things and then initiating an action, such as locking a door, changing equipment settings, applying the brakes on a train, zooming a video camera, opening a valve in response to a pressure reading, creating a bar chart, or sending an alert to a technician to make a preventive repair.

The defining characteristic of fog computing are as follows:

- **Contextual location awareness and low latency:** The fog node sits as close to the IoT endpoint as possible to deliver distributed computing.
- **Geographic distribution:** In sharp contrast to the more centralized cloud, the services and applications targeted by the fog nodes demand widely distributed deployments.
- **Deployment near IoT endpoints:** Fog nodes are typically deployed in the presence of a large number of IoT endpoints. For example, typical metering deployments often see 3000 to 4000 nodes per gateway router, which also functions as the fog computing node.
- **Wireless communication between the fog and the IoT endpoint:** Although it is possible to connect wired nodes, the advantages of fog are greatest when dealing with a large number of endpoints, and wireless access is the easiest way to achieve such scale.
- **Use for real-time interactions:** Important fog applications involve real-time interactions rather than batch processing. Preprocessing of data in the fog nodes allows upper-layer applications to perform batch processing on a subset of the data.

Edge Computing

Fog computing solutions are being adopted by many industries, and efforts to develop distributed applications and analytics tools are being introduced at an accelerating pace. The natural place for a fog node is in the network device that sits closest to the IoT endpoints, and these nodes are typically spread throughout an IoT network

Note

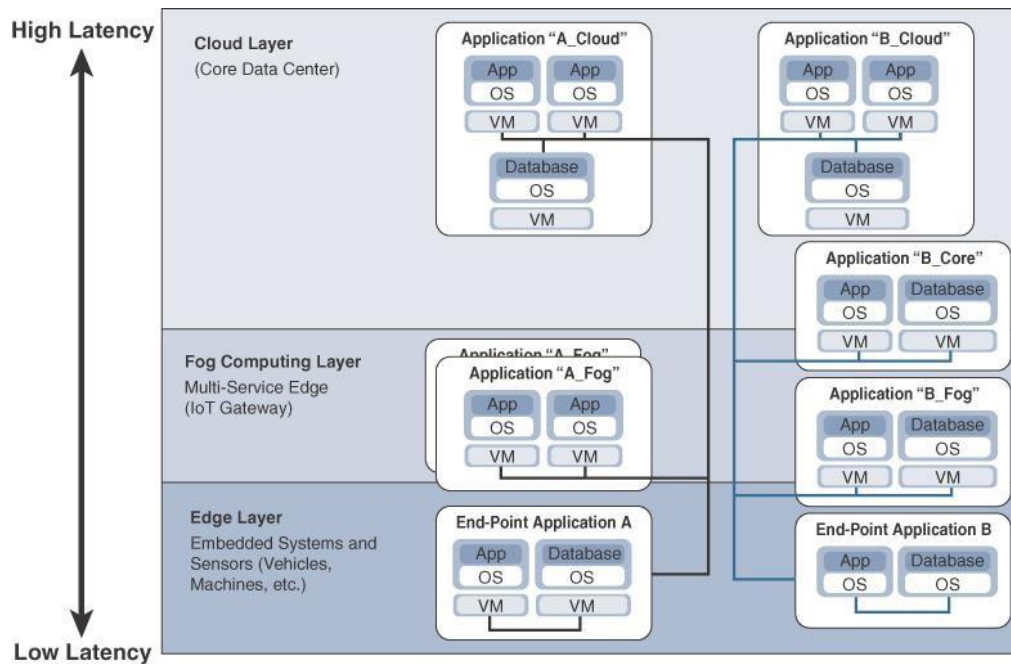
Edge computing is also sometimes called “mist” computing. If clouds exist in the sky, and fog sits near the ground, then mist is what actually sits on the ground. Thus, the concept of mist is to extend fog to the furthest point possible, right into the IoT endpoint device itself.

The Hierarchy of Edge, Fog, and Cloud

It is important to stress that edge or fog computing in no way replaces the cloud. Rather, they complement each other, and many use cases actually require strong cooperation between layers. In the same way that lower courts do not replace the supreme court of a

Internet of Things

country, edge and fog computing layers simply act as a first line of defense for filtering, analyzing, and otherwise managing data endpoints. This saves the cloud from being queried by each and every node for each event.



Distributed and Data Management Across an IoT System

From an architectural standpoint, fog nodes closest to the network edge receive the data from IoT devices. The fog IoT application then directs different types of data to the optimal place for analysis:

- The most time-sensitive data is analyzed on the edge or fog node closest to the things generating the data.
- Data that can wait seconds or minutes for action is passed along to an aggregation node for analysis and action.
- Data that is less time sensitive is sent to the cloud for historical analysis, big data analytics, and long-term storage. For example, each of thousands or hundreds of thousands of fog nodes might send periodic summaries of data to the cloud for historical analysis and storage.

In summary, when architecting an IoT network, you should consider the amount of data to be analyzed and the time sensitivity of this data. Understanding these factors will help you decide whether cloud computing is enough or whether edge or fog computing would improve your system efficiency. Fog computing accelerates awareness and response to events by eliminating a round trip to the cloud for analysis. It avoids the need for costly bandwidth additions by offloading gigabytes of network traffic from the core network. It also protects sensitive IoT data by analyzing it inside company walls.

MODULE 2

Smart Objects: The “Things” in IoT

3.5 Connecting Smart Objects

IoT devices and sensors must be connected to the network for their data to be utilized. In addition to the wide range of sensors, actuators, and smart objects that make up IoT, there are also a number of different protocols used to connect them. Here the characteristics and communications criteria that are important for the technologies that smart objects employ for their connectivity, along with a deeper dive into some of the major technologies being deployed today.

3.5.1 Communications Criteria

In the world of connecting “things,” a large number of wired and wireless access technologies are available or under development. Before reviewing some of these access technologies, it is important to talk about the criteria to use in evaluating them for various use cases and system solutions.

Wireless communication is prevalent in the world of smart object connectivity, mainly because it eases deployment and allows smart objects to be mobile, changing location without losing connectivity. The following sections take this into account as they discuss various criteria. In addition, wired connectivity considerations are mentioned when applicable.

3.5.1.1 Range

How far does the signal need to be propagated? That is, what will be the area of coverage for a selected wireless technology? Should indoor versus outdoor deployments be differentiated? Very often, these are the questions asked when discussing wired and wireless access technologies. The simplest approach to answering these types of questions is to categorize these technologies as shown in Figure 4-1, breaking them down into the following ranges:

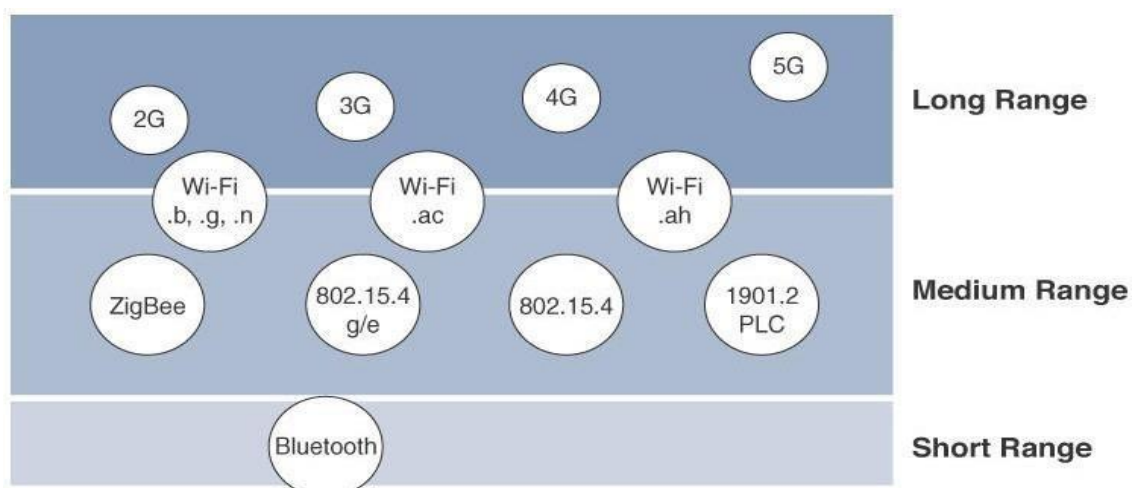


Figure 4-1 Wireless Access Landscape

- **Short range:** The classical wired example is a serial cable. Wireless short-range technologies are often considered as an alternative to a serial cable, supporting tens of meters of maximum distance between two devices. Examples of short-range wireless technologies are IEEE

802.15.1 Bluetooth and IEEE 802.15.7 Visible Light Communications (VLC). These short-range communication methods are found in only a minority of IoT installations. In some cases, they are not mature enough for production deployment.

- **Medium range:** This range is the main category of IoT access technologies. In the range of tens to hundreds of meters, many specifications and implementations are available. The maximum distance is generally less than 1 mile between two devices, although RF technologies do not have real maximum distances defined, as long as the radio signal is transmitted and received in the scope of the applicable specification. Examples of medium-range wireless technologies include IEEE 802.11 Wi-Fi, IEEE 802.15.4, and 802.15.4g WPAN. Wired technologies such as IEEE 802.3 Ethernet and IEEE 1901.2 Narrowband Power Line Communications (PLC) may also be classified as medium range, depending on their physical media characteristics.
- **Long range:** Distances greater than 1 mile between two devices require long-range technologies. Wireless examples are cellular (2G, 3G, 4G) and some applications of outdoor IEEE 802.11 Wi-Fi and Low-Power Wide-Area (LPWA) technologies. LPWA communications have the ability to communicate over a large area without consuming much power. These technologies are therefore ideal for battery-powered IoT sensors. Found mainly in industrial networks, IEEE 802.3 over optical fiber and IEEE 1901 Broadband Power Line Communications are classified as long range but are not really considered IoT access technologies.

3.5.1.2 Frequency Bands

Radio spectrum is regulated by countries and/or organizations, such as the International Telecommunication Union (ITU) and the Federal Communications Commission (FCC). These groups define the regulations and transmission requirements for various frequency bands. For example, portions of the spectrum are allocated to types of telecommunications such as radio, television, military, and so on.

Focusing on IoT access technologies, the frequency bands leveraged by wireless communications are split between licensed and unlicensed bands. Licensed spectrum is generally applicable to IoT long-range access technologies and allocated to communications infrastructures deployed by services providers, public services (for example, first responders, military), broadcasters, and utilities.

An important consideration for IoT access infrastructures that wish to utilize licensed spectrum is that users must subscribe to services when connecting their IoT devices. This adds more complexity to a deployment involving large numbers of sensors and other IoT devices, but in exchange for the subscription fee, the network operator can guarantee the exclusivity of the frequency usage over the target area and can therefore sell a better guarantee of service.

The ITU has also defined unlicensed spectrum for the industrial, scientific, and medical (ISM) portions of the radio bands. These frequencies are used in many communications technologies for short-range devices (SRDs). Unlicensed means that no guarantees or protections are offered in the ISM bands for device communications. For IoT access, these are the most well-known ISM bands:

- 2.4 GHz band as used by IEEE 802.11b/g/n Wi-Fi

- IEEE 802.15.1 Bluetooth
- IEEE 802.15.4 WPAN

An unlicensed band, such as those in the ISM range of frequencies, is not unregulated. National and regional regulations exist for each of the allocated frequency bands (much as with the licensed bands). These regulations mandate device compliance on parameters such as transmit power, duty cycle and dwell time, channel bandwidth, and channel hopping.

Unlicensed spectrum is usually simpler to deploy than licensed because it does not require a service provider. However, it can suffer from more interference because other devices may be competing for the same frequency in a specific area. This becomes a key element in decisions for IoT deployments.

Should an IoT infrastructure utilize unlicensed spectrum available for private networks or licensed frequencies that are dependent on a service provider? Various LPWA technologies are taking on a greater importance when it comes to answering this question. In addition to meeting low power requirements, LPWA communications are able to cover long distances that in the past required the licensed bands offered by service providers for cellular devices.

3.5.1.3 Power Consumption

While the definition of IoT device is very broad, there is a clear delineation between powered nodes and battery-powered nodes. A powered node has a direct connection to a power source, and communications are usually not limited by power consumption criteria. However, ease of deployment of powered nodes is limited by the availability of a power source, which makes mobility more complex.

Battery-powered nodes bring much more flexibility to IoT devices. These nodes are often classified by the required lifetimes of their batteries. Does a node need 10 to 15 years of battery life, such as on water or gas meters? Or is a 5- to 7-year battery life sufficient for devices such as smart parking sensors? Their batteries can be changed or the devices replaced when a street gets resurfaced. For devices under regular maintenance, a battery life of 2 to 3 years is an option.

IoT wireless access technologies must address the needs of low power consumption and connectivity for battery-powered nodes. This has led to the evolution of a new wireless environment known as Low-Power Wide-Area (LPWA). Obviously, it is possible to run just about any wireless technology on batteries. However, in reality, no operational deployment will be acceptable if hundreds of batteries must be changed every month.

Wired IoT access technologies consisting of powered nodes are not exempt from power optimization. In the case of deployment of smart meters over PLC, the radio interface on meters can't consume 5 to 10 watts of power, or this will add up to a 20-million-meter deployment consuming 100 to 200 megawatts of energy for communications.

3.5.1.4 Topology

Among the access technologies available for connecting IoT devices, three main topology schemes are

dominant: star, mesh, and peer-to-peer. For long-range and short-range technologies, a star topology is prevalent, as seen with cellular, LPWA, and Bluetooth networks. Star topologies utilize a single central base station or controller to allow communications with endpoints.

For medium-range technologies, a star, peer-to-peer, or mesh topology is common, as shown in Figure 4-2. Peer-to-peer topologies allow any device to communicate with any other device as long as they are in range of each other. Obviously, peer-to-peer topologies rely on multiple full-function devices. Peer-to-peer topologies enable more complex formations, such as a mesh networking topology.

A mesh topology helps cope with low transmit power, searching to reach a greater overall distance, and coverage by having intermediate nodes relaying traffic for other nodes. Mesh topology requires the implementation of a Layer 2 forwarding protocol known as mesh-under or a Layer 3 forwarding protocol referred to as mesh-over on each intermediate node. An intermediate node or full-function device (FFD) is simply a node that interconnects other nodes. A node that doesn't interconnect or relay the traffic of other nodes is known as a leaf node, or reduced-function device (RFD).

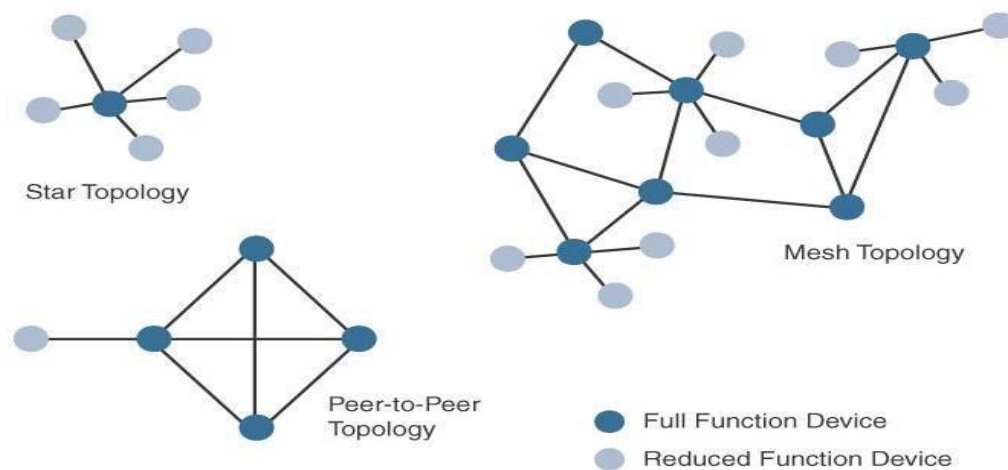


Figure 4-2 Star, Peer-to-Peer, and Mesh Topologies

3.5 Constrained Devices

The Internet Engineering Task Force (IETF) acknowledges in RFC 7228 that different categories of IoT devices are deployed. While categorizing the class of IoT nodes is a perilous exercise, with computing, memory, storage, power, and networking continuously evolving and improving, RFC 7228 gives some definitions of constrained nodes. These definitions help differentiate constrained nodes from unconstrained nodes, such as servers, desktop or laptop computers, and powerful mobile devices such as smart phones. According to RFC 7228, constrained nodes can be broken down into the classes defined in Table 4-1.

Class	Definition
Class 0	This class of nodes is severely constrained, with less than 10 KB of memory and less than 100 KB of Flash processing and storage capability. These nodes are typically battery powered. They do not have the resources required to directly implement an IP stack and associated security mechanisms. An example of a Class 0 node is a push button that sends 1 byte of information when changing its status. This class is particularly well suited to leveraging new unlicensed LPWA wireless technology.
Class 1	While greater than Class 0, the processing and code space characteristics (approximately 10 KB RAM and approximately 100 KB Flash) of Class 1 are still lower than expected for a complete IP stack implementation. They cannot easily communicate with nodes employing a full IP stack. However, these nodes can implement an optimized stack specifically designed for constrained nodes, such as Constrained Application Protocol (CoAP). This allows Class 1 nodes to engage in meaningful conversations with the network without the help of a gateway, and provides support for the necessary security functions. Environmental sensors are an example of Class 1 nodes.
Class 2	Class 2 nodes are characterized by running full implementations of an IP stack on embedded devices. They contain more than 50 KB of memory and 250 KB of Flash, so they can be fully integrated in IP networks. A smart power meter is an example of a Class 2 node.

Table 4-1 *Classes of Constrained Nodes, as Defined by RFC 7228*

3.6.1 Constrained-Node Networks

While several of the IoT access technologies, such as Wi-Fi and cellular, are applicable to laptops, smart phones, and some IoT devices, some IoT access technologies are more suited to specifically connect constrained nodes. Typical examples are IEEE 802.15.4 and 802.15.4g RF, IEEE 1901.2a PLC, LPWA, and IEEE 802.11ah access technologies.

Constrained-node networks are often referred to as **Low-power and Lossy Networks (LLNs)**. Low-power in the context of LLNs refers to the fact that nodes must cope with the requirements from powered and battery-powered constrained nodes. Lossy networks indicates that network performance may suffer from interference and variability due to harsh radio environments. Layer 1 and Layer 2 protocols that can be used for constrained-node networks must be evaluated in the context of the following characteristics for use-case applicability: data rate and throughput, latency and determinism, and overhead and payload.

1. Data Rate and Throughput

The data rates available from IoT access technologies range from 100 bps with protocols such as Sigfox to tens of megabits per second with technologies such as LTE and IEEE 802.11ac. However, the actual throughput is less—sometimes much less—than the data rate. Therefore, understanding the bandwidth requirements of a particular technology, its applicability to given use cases, the capacity

planning rules, and the expected real throughput are important for proper network design and successful production deployment.

Technologies not particularly designed for IoT, such as cellular and Wi-Fi, match up well to IoT applications with high bandwidth requirements. For example, nodes involved with video analytics have a need for high data rates. These nodes are found in retail, airport, and smart cities environments for detecting events and driving actions. Because these types of IoT endpoints are not constrained in terms of computing or network bandwidth, the design guidelines tend to focus on application requirements, such as latency and determinism.

Short-range technologies can also provide medium to high data rates that have enough throughput to connect a few endpoints. For example, Bluetooth sensors that are now appearing on connected wearables fall into this category. The IoT access technologies developed for constrained nodes are optimized for low power consumption, but they are also limited in terms of data rate, which depends on the selected frequency band, and throughput.

A discussion of data rate and bandwidth in LLNs must include a look at real throughput, or “goodput,” as seen by the application. While it may not be important for constrained nodes that send only one message a day, real throughput is often very important for constrained devices implementing an IP stack. In this case, throughput is a lower percentage of the data rate, even if the node gets the full constrained network at a given time.

Another characteristic of IoT devices is that a majority of them initiate the communication. Upstream traffic toward an application server is usually more common than downstream traffic from the application server. Understanding this behavior also helps when deploying an IoT access technology, such as cellular, that is asymmetrical because the upstream bandwidth must be considered a key parameter for profiling the network capacity.

Latency and Determinism

Much like throughput requirements, latency expectations of IoT applications should be known when selecting an access technology. This is particularly true for wireless networks, where packet loss and retransmissions due to interference, collisions, and noise are normal behaviors.

On constrained networks, latency may range from a few milliseconds to seconds, and applications and protocol stacks must cope with these wide-ranging values. For example, UDP at the transport layer is strongly recommended for IP endpoints communicating over LLNs. In the case of mesh topologies, if communications are needed between two devices inside the mesh, the forwarding path may call for some routing optimization, which is available using the IPv6 RPL protocol.

Overhead and Payload

When considering constrained access network technologies, it is important to review the MAC payload size characteristics required by applications. The minimum IPv6 MTU size is expected to be

1280 bytes. Therefore, the fragmentation of the IPv6 payload has to be taken into account by link layer access protocols with smaller MTUs.

For example, the payload size for IEEE 802.15.4 is 127 bytes and requires an IPv6 payload with a minimum MTU of 1280 bytes to be fragmented. On the other hand, IEEE 802.15.4g enables payloads up to 2048 bytes, easing the support of the IPv6 minimum MTU of 1280 bytes.

Most LPWA technologies offer small payload sizes. These small payload sizes are defined to cope with the low data rate and time over the air or duty cycle requirements of IoT nodes and sensors. For example, payloads may be as little as 19 bytes using LoRaWAN technology or up to 250 bytes, depending on the adaptive data rate (ADR). While this doesn't preclude the use of an IPv6/6LoWPAN payload, as seen on some endpoint implementations, these types of protocols are better suited to Class 0 and 1 nodes, as defined in RFC 7228.

3.6 IoT Access Technologies

The technologies highlighted here are the ones that are seen as having market and/or mind share. Therefore, it is necessary to have a basic familiarity with them as they are fundamental to many IoT conversations. For each of the IoT access technologies discussed here, a common information set is being provided. Particularly, the following topics are addressed for each IoT access technology:

- **Standardization and alliances:** The standards bodies that maintain the protocols for a technology
- **Physical layer:** The wired or wireless methods and relevant frequencies
- **MAC layer:** Considerations at the Media Access Control (MAC) layer, which bridges the physical layer with data link control
- **Topology:** The topologies supported by the technology
- **Security:** Security aspects of the technology
- **Competitive technologies:** Other technologies that are similar and may be suitable alternatives to the given technology

3.6.1 IEEE 802.15.4

IEEE 802.15.4 is a wireless access technology for low-cost and low-data-rate devices that are powered or run on batteries. In addition to being low cost and offering a reasonable battery life, this access technology enables easy installation using a compact protocol stack while remaining both simple and flexible. Several network communication stacks, including deterministic ones, and profiles leverage this technology to address a wide range of IoT use cases in both the consumer and business markets. IEEE 802.15.4 is commonly found in the following types of deployments:

- Home and building automation
- Automotive networks
- Industrial wireless sensor networks
- Interactive toys and remote controls

Criticisms of IEEE 802.15.4 often focus on its MAC reliability, unbounded latency, and susceptibility to interference and multipath fading. The negatives around reliability and latency often have to do with the Collision Sense Multiple Access/Collision Avoidance (CSMA/CA) algorithm. CSMA/CA is an access method in which a device “listens” to make sure no other devices are transmitting before starting its own transmission.

Standardization and Alliances

IEEE 802.15.4 or IEEE 802.15 Task Group 4 defines low-data-rate PHY and MAC layer specifications for wireless personal area networks (WPAN). This standard has evolved over the years and is a well-known solution for low-complexity wireless devices with low data rates that need many months or even years of battery life.

While there is no alliance or promotion body for IEEE 802.15.4 per se, the IEEE 802.15.4 PHY and MAC layers are the foundations for several networking protocol stacks. These protocol stacks make use of 802.15.4 at the physical and link layer levels, but the upper layers are different. These protocol stacks are promoted separately through various organizations and often commercialized. Some of the most well-known protocol stacks based on 802.15.4 are highlighted in Table 4-2.

Protocol	Description
ZigBee	Promoted through the ZigBee Alliance, ZigBee defines upper-layer components (network through application) as well as application profiles. Common profiles include building automation, home automation, and healthcare. ZigBee also defines device object functions, such as device role, device discovery, network join, and security. For more information on ZigBee, see the ZigBee Alliance webpage, at www.zigbee.org . ZigBee is also discussed in more detail later in the next Section.
6LoWPAN	6LoWPAN is an IPv6 adaptation layer defined by the IETF 6LoWPAN working group that describes how to transport IPv6 packets over IEEE 802.15.4 layers. RFCs document header compression and IPv6 enhancements to cope with the specific details of IEEE 802.15.4. (For more information on 6LoWPAN, see Chapter 5.)

ZigBee IP	An evolution of the ZigBee protocol stack, ZigBee IP adopts the 6LoWPAN adaptation layer, IPv6 network layer, and RPL routing protocol. In addition, it offers improvements to IP security. ZigBee IP is discussed in more detail later in this chapter.
ISA100.11a	ISA100.11a is developed by the International Society of Automation (ISA) as “Wireless Systems for Industrial Automation: Process Control and Related Applications.” It is based on IEEE 802.15.4-2006, and specifications were published in 2010 and then as IEC 62734. The network and transport layers are based on IETF 6LoWPAN, IPv6, and UDP standards.
WirelessHART	WirelessHART, promoted by the HART Communication Foundation, is a protocol stack that offers a time-synchronized, self-organizing, and self-healing mesh architecture, leveraging IEEE 802.15.4-2006 over the 2.4 GHz frequency band. A good white paper on WirelessHART can be found at http://www.emerson.com/resource/blob/system-engineering-guidelines-iec-62591-wirelesshart--data-79900.pdf
Thread	Constructed on top of IETF 6LoWPAN/IPv6, Thread is a protocol stack for a secure and reliable mesh network to connect and control products in the home. Specifications are defined and published by the Thread Group at www.threadgroup.org .

Table 4-2 Protocol Stacks Utilizing IEEE 802.15.4

Because of its relatively long history compared to the others, ZigBee is one of the most well-known protocols listed in Table 4-2. In addition, ZigBee has continued to evolve over time as evidenced by the release of Zigbee IP and is representative of how IEEE 802.15.4 can be leveraged at the PHY and MAC layers, independent of the protocol layers above. For these reasons, both Zigbee and Zigbee IP are discussed in more detail in the following sections.

ZigBee

Based on the idea of ZigBee-style networks in the late 1990s, the first ZigBee specification was ratified in 2004, shortly after the release of the IEEE 802.15.4 specification the previous year. While not released as a typical standard, like an RFC, ZigBee still had industry support from more than 100 companies upon its initial publication. This industry support has grown to more than 400 companies that are members of the ZigBee Alliance. Similar to the Wi-Fi Alliance, the Zigbee Alliance is an industry group formed to certify interoperability between vendors and it is committed to driving and evolving ZigBee as an IoT solution for interconnecting smart objects.

ZigBee solutions are aimed at smart objects and sensors that have low bandwidth and low power needs. Furthermore, products that are ZigBee compliant and certified by the ZigBee Alliance should interoperate even though different vendors may manufacture them.

The main areas where ZigBee is the most well-known include automation for commercial, retail, and home applications and smart energy. In the industrial and commercial automation space, ZigBee-based devices can handle various functions, from measuring temperature and humidity to tracking assets. For home automation, ZigBee can control lighting, thermostats, and security functions.

ZigBee Smart Energy brings together a variety of interoperable products, such as smart meters, that can monitor and control the use and delivery of utilities, such as electricity and water.

The traditional ZigBee stack is illustrated in Figure 4-3. As mentioned previously, ZigBee utilizes the IEEE 802.15.4 standard at the lower PHY and MAC layers. ZigBee specifies the network and security layer and application support layer that sit on top of the lower layers.

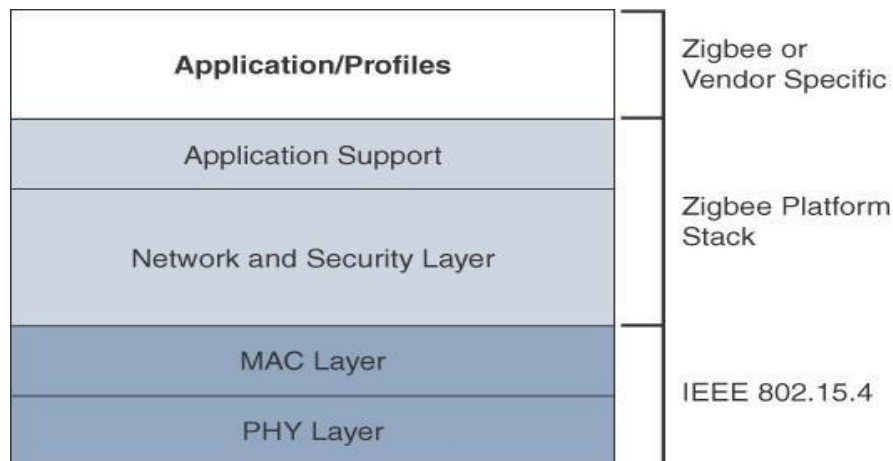


Figure 4-3 High-Level ZigBee Protocol Stack

The ZigBee network and security layer provides mechanisms for network startup, configuration, routing, and securing communications. This includes calculating routing paths in what is often a changing topology, discovering neighbors, and managing the routing tables as devices join for the first time. The network layer is also responsible for forming the appropriate topology, which is often a mesh but could be a star or tree as well. From a security perspective, ZigBee utilizes 802.15.4 for security at the MAC layer, using the Advanced Encryption Standard (AES) with a 128-bit key and also provides security at the network and application layers.

ZigBee IP

With the introduction of ZigBee IP, the support of IEEE 802.15.4 continues, but the IP and TCP/UDP protocols and various other open standards are now supported at the network and transport layers. The ZigBee-specific layers are now found only at the top of the protocol stack for the applications. ZigBee IP was created to embrace the open standards coming from the IETF's work on LLNs, such as IPv6, 6LoWPAN, and RPL. They provide for low-bandwidth, low-power, and cost-effective communications when connecting smart objects.

ZigBee IP is a critical part of the Smart Energy (SE) Profile 2.0 specification from the ZigBee Alliance. SE 2.0 is aimed at smart metering and residential energy management systems. In fact, ZigBee IP was designed specifically for SE 2.0 but it is not limited to this use case. Any other applications that need a standards-based IoT stack can utilize Zigbee IP. The ZigBee IP stack is shown in Figure 4-4.

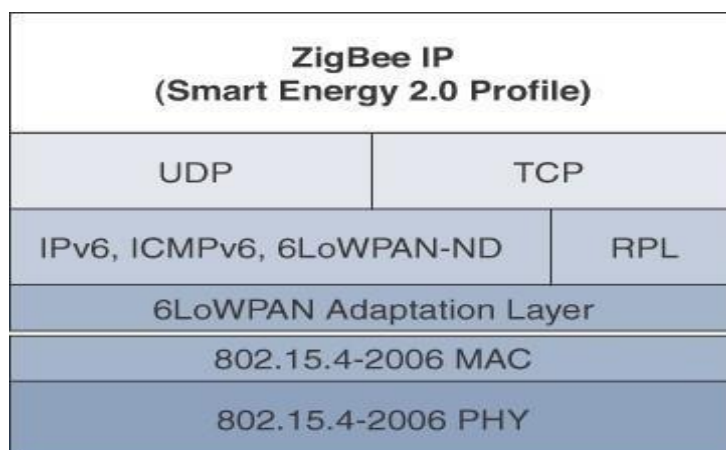


Figure 4-4 ZigBee IP Protocol Stack

ZigBee IP supports 6LoWPAN as an adaptation layer. The 6LoWPAN mesh addressing header is not required as ZigBee IP utilizes the mesh-over or route-over method for forwarding packets. ZigBee IP requires the support of 6LoWPAN's fragmentation and header compression schemes. At the network layer, all ZigBee IP nodes support IPv6, ICMPv6, and 6LoWPAN Neighbor Discovery (ND), and utilize RPL for the routing of packets across the mesh network. Both TCP and UDP are also supported, to provide both connection-oriented and connectionless service.

Physical Layer

The 802.15.4 standard supports an extensive number of PHY options that range from 2.4 GHz to sub-GHz frequencies in ISM bands. original IEEE 802.15.4-2003 standard specified only three PHY options based on **Direct Sequence Spread Spectrum (DSSS)** modulation. DSSS is a modulation technique in which a signal is intentionally spread in the frequency domain, resulting in greater bandwidth. The original physical layer transmission options were as follows:

- 2.4 GHz, 16 channels, with a data rate of 250 kbps
- 915 MHz, 10 channels, with a data rate of 40 kbps
- 868 MHz, 1 channel, with a data rate of 20 kbps

Note that only the 2.4 GHz band operates worldwide. The 915 MHz band operates mainly in North and South America, and the 868 MHz frequencies are used in Europe, the Middle East, and Africa. IEEE 802.15.4-2006, 802.15.4-2011, and IEEE 802.15.4-2015 introduced additional PHY communication options, including the following:

- **OQPHY:** This is DSSS PHY, employing offset quadrature phase-shift keying (OQPSK) modulation. OQPSK is a modulation technique that uses four unique bit values that are signaled by phase changes. An offset function that is present during phase shifts allows data to be transmitted more reliably.
- **BPSK PHY:** This is DSSS PHY, employing binary phase-shift keying (BPSK) modulation. BPSK specifies two unique phase shifts as its data encoding scheme.
- **ASK PHY:** This is parallel sequence spread spectrum (PSSS) PHY, employing amplitude shift keying (ASK) and BPSK modulation. PSSS is an advanced encoding scheme that offers increased range, throughput, data rates, and signal integrity compared to DSSS. ASK uses amplitude shifts instead of phase shifts to signal different bit values.

Figure 4-5 shows the frame for the 802.15.4 physical layer. The synchronization header for this frame is composed of the Preamble and the Start of Frame Delimiter fields. The Preamble field is a 32-bit 4-byte (for parallel construction) pattern that identifies the start of the frame and is used to synchronize the data transmission. The Start of Frame Delimiter field informs the receiver that frame contents start immediately after this byte.

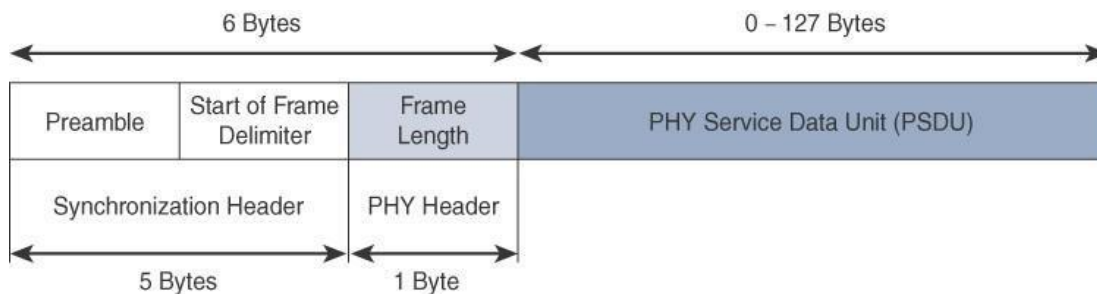


Figure 4-5 IEEE 802.15.4 PHY Format

The PHY Header portion of the PHY frame shown in Figure 4-5 is simply a frame length value. It lets the receiver know how much total data to expect in the PHY service data unit (PSDU) portion of the 802.15.4 PHY. The PSDU is the data field or payload.

MAC Layer

The IEEE 802.15.4 MAC layer manages access to the PHY channel by defining how devices in the same area will share the frequencies allocated. At this layer, the scheduling and routing of data frames are also coordinated. The 802.15.4 MAC layer performs the following tasks:

- Network beaconing for devices acting as coordinators (New devices use beacons to join an 802.15.4 network)
- PAN association and disassociation by a device
- Device security
- Reliable link communications between two peer MAC entities

The MAC layer achieves these tasks by using various predefined frame types. In fact, four types of MAC frames are specified in 802.15.4:

- **Data frame:** Handles all transfers of data
- **Beacon frame:** Used in the transmission of beacons from a PAN coordinator
- **Acknowledgement frame:** Confirms the successful reception of a frame
- **MAC command frame:** Responsible for control communication between devices

Each of these four 802.15.4 MAC frame types follows the frame format shown in Figure 4-6. In Figure 4-6, notice that the MAC frame is carried as the PHY payload. The 802.15.4 MAC frame can be broken down into the MAC Header, MAC Payload, and MAC Footer fields.

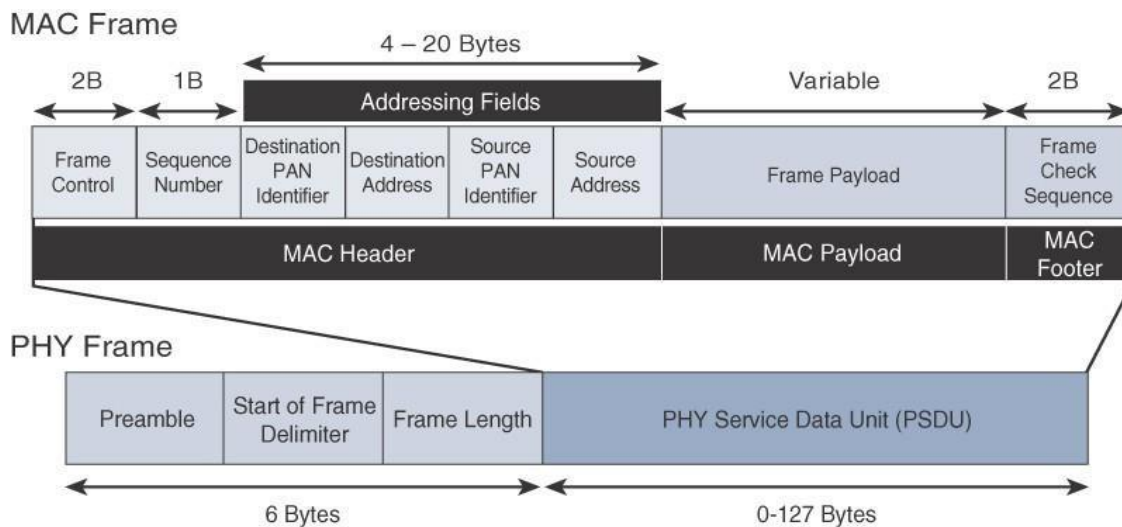


Figure 4-6 IEEE 802.15.4 MAC Format

- The MAC Header field is composed of the Frame Control, Sequence Number and the Addressing fields. The Frame Control field defines attributes such as frame type, addressing modes, and other control flags. The Sequence Number field indicates the sequence identifier for the frame. The Addressing field specifies the Source and Destination PAN Identifier fields as well as the Source and Destination Address fields.
- The MAC Payload field varies by individual frame type. For example, beacon frames have specific fields and payloads related to beacons, while MAC command frames have different fields present.
- The MAC Footer field is nothing more than a frame check sequence (FCS). An FCS is a calculation based on the data in the frame that is used by the receiving side to confirm the integrity of the data in the frame.

Topology

IEEE 802.15.4-based networks can be built as star, peer-to-peer, or mesh topologies. Mesh networks tie together many nodes. This allows nodes that would be out of range if trying to communicate directly to leverage intermediary nodes to transfer communications. Please note that every 802.15.4 PAN should be set up with a unique ID. All the nodes in the same 802.15.4 network should use the same PAN ID. Figure 4-7 shows an example of an 802.15.4 mesh network with a PAN ID of 1.

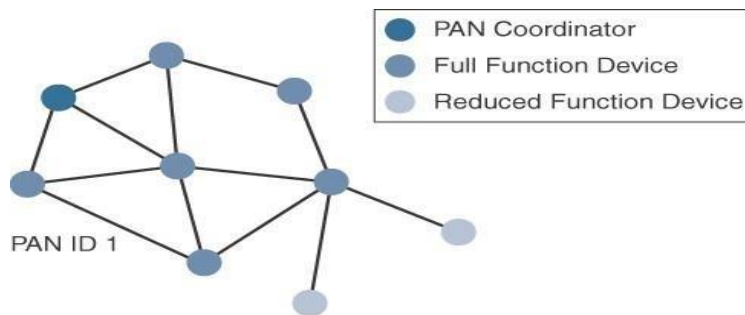


Figure 4-7 802.15.4 Sample Mesh Network Topology

A minimum of one FFD acting as a PAN coordinator is required to deliver services that allow other devices to associate and form a cell or PAN. Notice in Figure 4-7 that a single PAN coordinator is identified for PAN ID 1. FFD devices can communicate with any other devices, whereas RFD devices can communicate only with FFD devices.

Security

The IEEE 802.15.4 specification uses Advanced Encryption Standard (AES) with a 128-bit key length as the base encryption algorithm for securing its data. Established by the US National Institute of Standards and Technology in 2001, AES is a block cipher, which means it operates on fixed-size blocks of data. The use of AES by the US government and its widespread adoption in the private sector has helped it become one of the most popular algorithms used in symmetric key cryptography. (A symmetric key means that the same key is used for both the encryption and decryption of the data.) In addition to encrypting the data, AES in 802.15.4 also validates the data that is sent. This is accomplished by a message integrity code (MIC), which is calculated for the entire frame using the same AES key that is used for encryption.

Enabling these security features for 802.15.4 changes the frame format slightly and consumes some of the payload. Using the Security Enabled field in the Frame Control portion of the 802.15.4 header is the first step to enabling AES encryption. This field is a single bit that is set to 1 for security. Once this bit is set, a field called the Auxiliary Security Header is created after the Source Address field, by stealing some bytes from the Payload field. Figure 4-8 shows the IEEE 802.15.4 frame format at a high level, with the Security Enabled bit set and the Auxiliary Security Header field present.

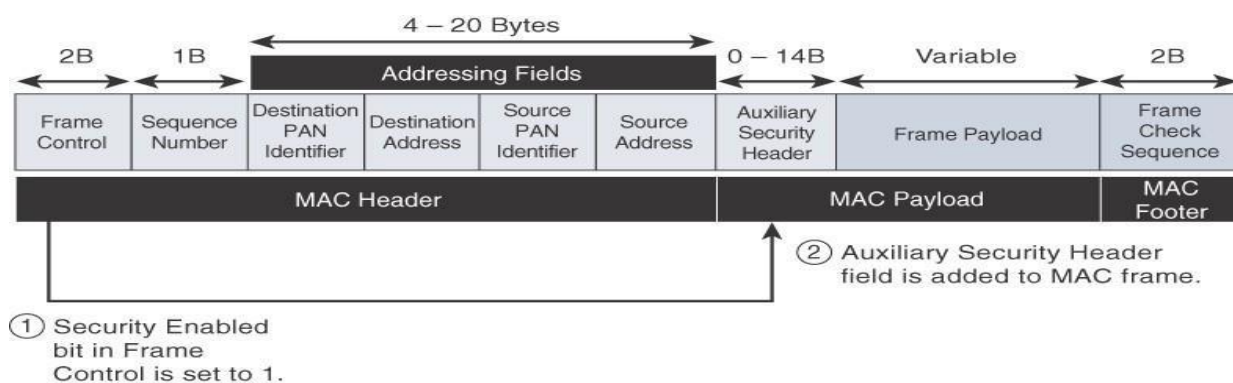


Figure 4-8 Frame Format with the Auxiliary Security Header Field for 802.15.4-2006 and Later Versions

Competitive Technologies

A competitive radio technology that is different in its PHY and MAC layers is DASH7. DASH7 was originally based on the ISO18000-7 standard and positioned for industrial communications, whereas IEEE 802.15.4 is more generic. Commonly employed in active radio frequency identification (RFID) implementations, DASH7 was used by US military forces for many years, mainly for logistics purposes. Active RFID utilizes radio waves generated by a battery-powered tag on an object to enable continuous tracking.

The current DASH7 technology offers low power consumption, a compact protocol stack, range up to 1 mile, and AES encryption. Frequencies of 433 MHz, 868 MHz, and 915 MHz have been defined, enabling data rates up to 166.667 kbps and a maximum payload of 256 bytes. DASH7 is promoted by the DASH7 Alliance, which has evolved the protocol from its active RFID niche into a wireless sensor network technology that is aimed at the commercial market.

3.6.2 IEEE 802.15.4g and 802.15.4e

The IEEE frequently makes amendments to the core 802.15.4 specification, before integrating them into the next revision of the core specification. When these amendments are made, a lowercase letter is appended. Two such examples of this are 802.15.4e-2012 and 802.15.4g-2012, both of which are especially relevant to the subject of IoT. Both of these amendments were integrated in IEEE 802.15.4-2015 but are often still referred to by their amendment names.

The IEEE 802.15.4e amendment of 802.15.4-2011 expands the MAC layer feature set to remedy the disadvantages associated with 802.15.4, including MAC reliability, unbounded latency, and multipath fading. In addition to making general enhancements to the MAC layer, IEEE 802.15.4e also made improvements to better cope with certain application domains, such as factory and process automation and smart grid.

IEEE 802.15.4g-2012 is also an amendment to the IEEE 802.15.4-2011 standard, and just like 802.15.4e-2012, it has been fully integrated into the core IEEE 802.15.4-2015 specification. The focus of this specification is the smart grid or, more specifically, smart utility network communication. 802.15.4g seeks to optimize large outdoor wireless mesh networks for field area networks (FANs). New PHY definitions are introduced, as well as some MAC modifications needed to support their implementation. This technology applies to IoT use cases such as the following:

- Distribution automation and industrial **supervisory control and data acquisition (SCADA)** environments for remote monitoring and control.
- Public lighting
- Image Environmental wireless sensors in smart cities
- Electrical vehicle charging stations
- Smart parking meters
- Microgrids
- Renewable energy

Standardization and Alliances

Because 802.15.4g-2012 and 802.15.4e-2012 are simply amendments to IEEE 802.15.4-2011, the same IEEE 802.15 Task Group 4 standards body authors, maintains, and integrates them into the next release of the core specification. However, the additional capabilities and options provided by 802.15.4g-2012 and 802.15.4e-2012 led to additional difficulty in achieving the interoperability between devices and mixed vendors that users requested.

To guarantee interoperability, the Wi-SUN Alliance was formed. (SUN stands for smart utility network.) This organization is not a standards body but is instead an industry alliance that defines communication profiles for smart utility and related networks. The Wi-SUN Alliance performs the same function as the Wi-Fi Alliance and WiMAX Forum. Each of these organizations has an associated standards body as well as a commercial name, as shown in Table 4-3.

Commercial Name/Trademark	Industry Organization	Standards Body
Wi-Fi	Wi-Fi Alliance	IEEE 802.11 Wireless LAN
WiMAX	WiMAX Forum	IEEE 802.16 Wireless MAN
Wi-SUN	Wi-SUN Alliance	IEEE 802.15.4g Wireless SUN

Table 4-3 Industry Alliances for Some Common IEEE Standards

Physical Layer

In IEEE 802.15.4g-2012, the original IEEE 802.15.4 maximum PSDU or payload size of 127 bytes was increased for the SUN PHY to 2047 bytes. This provides a better match for the greater packet sizes found in many upper-layer protocols. For example, the default IPv6 MTU setting is 1280 bytes. Fragmentation is no longer necessary at Layer 2 when IPv6 packets are transmitted over IEEE 802.15.4g MAC frames. Also, the error protection was improved in IEEE 802.15.4g by evolving the CRC from 16 to 32 bits.

The SUN PHY, as described in IEEE 802.15.4g-2012, supports multiple data rates in bands ranging from 169 MHz to 2.4 GHz. These bands are covered in the unlicensed ISM frequency spectrum specified by various countries and regions. Within these bands, data must be modulated onto the frequency using at least one of the following PHY mechanisms to be IEEE 802.15.4g compliant:

- **Multi-Rate and Multi-Regional Frequency Shift Keying (MR-FSK):** Offers good transmit power efficiency due to the constant envelope of the transmit signal
- **Multi-Rate and Multi-Regional Orthogonal Frequency Division Multiplexing (MR-OFDM):** Provides higher data rates but may be too complex for low-cost and low-power devices
- **Multi-Rate and Multi-Regional Offset Quadrature Phase-Shift Keying (MR-O-QPSK):** Shares the same characteristics of the IEEE 802.15.4-2006 O-QPSK PHY, making multi-mode systems more cost-effective and easier to design

Enhanced data rates and a greater number of channels for channel hopping are available, depending on the frequency bands and modulation. For example, for the 902–928 MHz ISM band that is used in the United States, MR-FSK provides 50, 150, or 200 kbps. MR-OFDM at this same frequency allows up to 800 kbps. Other frequencies provide their own settings.

MAC Layer

While the IEEE 802.15.4e-2012 amendment is not applicable to the PHY layer, it is pertinent to the MAC layer. This amendment enhances the MAC layer through various functions, which may be selectively enabled based on various implementations of the standard. In fact, if interoperability is a “must have,” then using profiles defined by organizations such as Wi-SUN is necessary. The following are some of the main enhancements to the MAC layer proposed by IEEE 802.15.4e-2012:

- **Time-Slotted Channel Hopping (TSCH):** TSCH is an IEEE 802.15.4e-2012 MAC operation mode that works to guarantee media access and channel diversity. Channel hopping, also known as frequency hopping, utilizes different channels for transmission at different times. TSCH divides time into fixed time periods, or “time slots,” which offer guaranteed bandwidth and predictable latency. In a time slot, one packet and its acknowledgement can be transmitted, increasing network capacity because multiple nodes can communicate in the same time slot, using different channels.
- **Information elements (IEs):** Information elements (IEs) allow for the exchange of information at the MAC layer in an extensible manner, either as header IEs (standardized) and/or payload IEs (private). Specified in a tag, length, value (TLV) format, the IE field allows frames to carry additional metadata to support MAC layer services. These services may include IEEE 802.15.9 key management, Wi-SUN 1.0 IEs to broadcast and unicast schedule timing information, and frequency hopping synchronization information for the 6TiSCH architecture.
- **Enhanced beacons (EBs):** EBs extend the flexibility of IEEE 802.15.4 beacons to allow the construction of application-specific beacon content. This is accomplished by including relevant IEs in EB frames. Some IEs that may be found in EBs include network metrics, frequency hopping broadcast schedule, and PAN information version.
- **Enhanced beacon requests (EBRs):** Like enhanced beacons, an enhanced beacon request (EBRs) also leverages IEs. The IEs in EBRs allow the sender to selectively specify the request of information. Beacon responses are then limited to what was requested in the EBR. For example, a device can query for a PAN that is allowing new devices to join or a PAN that supports a certain set of MAC/PHY capabilities.
- **Enhanced Acknowledgement:** The Enhanced Acknowledgement frame allows for the integration of a frame counter for the frame being acknowledged. This feature helps protect against certain attacks that occur when Acknowledgement frames are spoofed.

The 802.15.4e-2012 MAC amendment is quite often paired with the 802.15.4g-2012 PHY. Figure 4-9 details this frame format. Notice that the 802.15.4g-2012 PHY is similar to the 802.15.4 PHY in Figure 4-5. The main difference between the two is the payload size, with 802.15.4g supporting up to 2047 bytes and 802.15.4 supporting only 127 bytes.

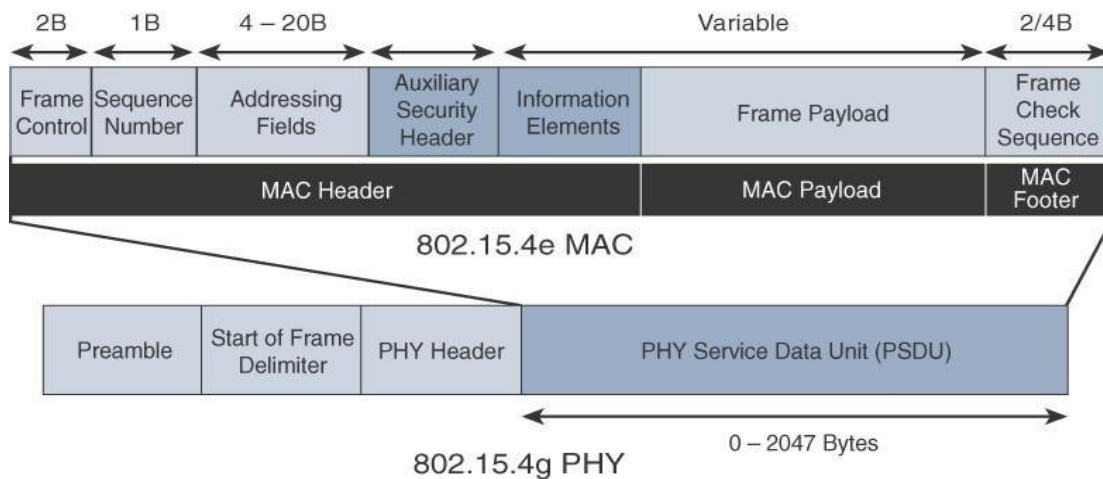


Figure 4-9 IEEE 802.15.4g/e MAC Frame Format

The 802.15.4e MAC is similar to the 802.15.4 MAC in Figure 4-6. The main changes shown in the IEEE 802.15.4e header in Figure 4-9 are the presence of the Auxiliary Security Header and Information Elements field.

- The Auxiliary Security header provides for the encryption of the data frame. This field is optionally supported in both 802.15.4e-2012 and 802.15.4, starting with the 802.15.4-2006 specification.
- The IE field contains one or more information elements that allow for additional information to be exchanged at the MAC layer.

Topology

Deployments of IEEE 802.15.4g-2012 are mostly based on a mesh topology. This is because a mesh topology is typically the best choice for use cases in the industrial and smart cities areas where 802.15.4g-2012 is applied. A mesh topology allows deployments to be done in urban or rural areas, expanding the distance between nodes that can relay the traffic of other nodes. Considering the use cases addressed by this technology, powered nodes have been the primary targets of implementations.

Security

Both IEEE 802.15.4g and 802.15.4e inherit their security attributes from the IEEE 802.15.4-2006 specification. Therefore, encryption is provided by AES, with a 128-bit key. In addition to the Auxiliary Security Header field initially defined in 802.15.4-2006, a secure acknowledgement and a secure Enhanced Beacon field complete the MAC layer security. Figure 4-10 shows a high-level overview of the security associated with an IEEE 802.15.4e MAC frame.

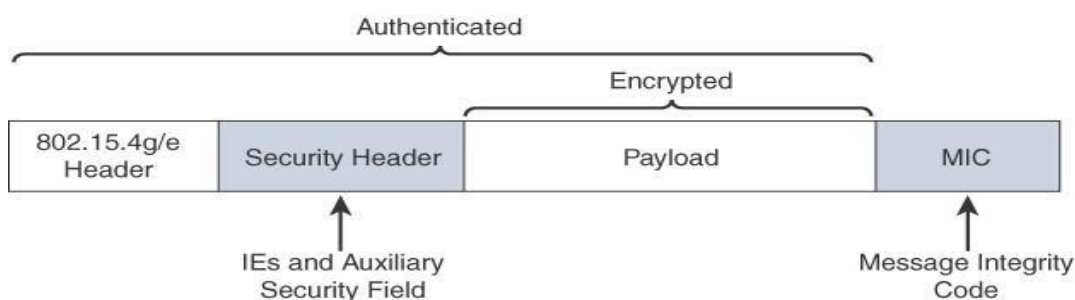


Figure 4-10 IEEE 802.15.4g/e MAC Layer Security

The full frame in Figure 4-10 gets authenticated through the MIC at the end of frame. The MIC is a unique value that is calculated based on the frame contents. The Security Header field denoted in Figure 4-10 is composed of the Auxiliary Security field and one or more Information Elements fields. Integration of the Information Elements fields allows for the adoption of additional security capabilities, such as the IEEE 802.15.9 Key Management Protocol (KMP) specification. KMP provides a means for establishing keys for robust datagram security. Without key management support, weak keys are often the result, leaving the security system open to attack.

Competitive Technologies

Competitive technologies to IEEE 802.15.4g and 802.15.4e parallel the technologies that also compete with IEEE 802.15.4, such as DASH7. In many ways, 802.15.4 and its various flavors of upper-layer protocols, as shown in Table 4-2, can be seen as competitors as well. IEEE 802.15.4 is well established and already deployed in many scenarios, mostly indoors.

3.6.3 IEEE 1901.2a

While most of the constrained network technologies relate to wireless, IEEE 1901.2a-2013 is a wired technology that is an update to the original IEEE 1901.2 specification. This is a standard for Narrowband Power Line Communication (NB-PLC). NB-PLC is a narrowband spectrum for low power, long range, and resistance to interference over the same wires that carry electric power. NB-PLC is often found in use cases such as the following

- **Smart metering:** NB-PLC can be used to automate the reading of utility meters, such as electric, gas, and water meters. This is true particularly in Europe, where PLC is the preferred technology for utilities deploying smart meter solutions.
- **Distribution automation:** NB-PLC can be used for distribution automation, which involves monitoring and controlling all the devices in the power grid.
- **Public lighting:** A common use for NB-PLC is with public lighting—the lights found in cities and along streets, highways, and public areas such as parks.
- **Electric vehicle charging stations:** NB-PLC can be used for electric vehicle charging stations, where the batteries of electric vehicles can be recharged.
- **Microgrids:** NB-PLC can be used for microgrids, local energy grids that can disconnect from the traditional grid and operate independently.
- **Renewable energy:** NB-PLC can be used in renewable energy applications, such as solar, wind power, hydroelectric, and geothermal heat.

All these use cases require a direct connection to the power grid. So it makes sense to transport IoT data across power grid connections that are already in place.

Standardization and Alliances

The first generations of NB-PLC implementations have generated a lot of interest from utilities in Europe but have often suffered from poor reliability, low throughput (in the range of a few hundred bits per second to a maximum of 2 kbps), lack of manageability, and poor interoperability. This has led several organizations (including standards bodies and alliance consortiums) to develop their own specifications for new generations of NB-PLC technologies. Most recent NB-PLC standards are based on orthogonal frequency-division multiplexing (OFDM). However, different standards from various vendors competing with one another have created a fragmented market. OFDM encodes digital data

on multiple carrier frequencies. This provides several parallel streams that suffer less from high frequency attenuation in copper wire and narrowband interference.

The HomePlug Alliance was one of the main industry organizations that drove the promotion and certification of PLC technologies, with IEEE 1901.2a being part of its HomePlug Netricity program. In 2016, the HomePlug Alliance made the decision to offer the alliance's broadband power line networking technology to a broader audience by making its technical specifications publicly available. It has also partnered with other alliances on continuing ongoing work. The HomePlug Alliance has struck a liaison agreement with the Wi-SUN Alliance with the goal of enabling hybrid smart grid networks that support both wireless and power line-wired connectivity. For more information on the HomePlug Alliance and Netricity, see www.homeplug.org.

Physical Layer

NB-PLC is defined for frequency bands from 3 to 500 kHz. Much as with wireless sub-GHz frequency bands, regional regulations and definitions apply to NB-PLC. The IEEE 1901.2 working group has integrated support for all world regions in order to develop a worldwide standard. Specifications include support for CENELEC A and B bands, US FCC-Low and FCC-above- CENELEC, and Japan ARIB bands. CENELEC is the French Comité Européen de Normalisation Électrotechnique, which in English translates to European Committee for Electrotechnical Standardization. This organization is responsible for standardization in the area of electrical engineering for Europe. The CENELEC A and B bands refer to 9–95 kHz and 95–125 kHz, respectively.

The FCC is the Federal Communications Commission, a US government organization that regulates interstate and international communications by radio, television, wire, satellite, and cable. The FCC-Low band encompasses 37.5–117.1875 kHz, and the FCC-above-CENELEC band is 154.6875–487.5 kHz. The FCC-above-CENELEC band may become the most useful frequency due to its higher throughput and reduced interference.

Figure 4-11 shows the various frequency bands for NB-PLC. Notice that the most well-known bands are regulated by CENELEC and the FCC, but the Japan Association of Radio Industries and Businesses (ARIB) band is also present. The two ARIB frequency bands are ARIB 1, 37.5–117.1875 kHz, and ARIB 2, 154.6875–403.125 kHz.

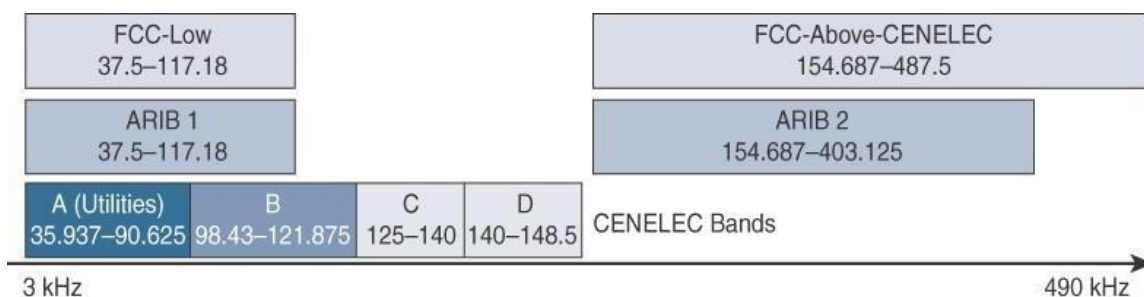


Figure 4-11 NB-PLC Frequency Bands

Based on OFDM, the IEEE 1901.2 specification leverages the best from other NB-PLC OFDM technologies that were developed previously. Therefore, IEEE 1901.2a supports the largest set of coding and enables both robustness and throughput. The standard includes tone maps and modulations,

such as robust modulation (ROBO), differential binary phase shift keying (DBPSK), differential quadrature phase shift keying (DQPSK), differential 8-point phase shift keying (D8PSK) for all bands, and optionally 16 quadrature amplitude modulation (16QAM) for some bands. ROBO mode transmits redundant information on multiple carriers, and DBPSK, DQPSK, and D8PSK are all variations of phase shift keying, where the phase of a signal is changed to signal a binary data transmission. ROBO utilizes QPSK modulation, and its throughput depends on the degree to which coding is repeated across streams. For example, standard ROBO uses a repetition of 4, and Super- ROBO utilizes a repetition of 6.

With IEEE 1901.2a, the data throughput rate has the ability to dynamically change, depending on the modulation type and tone map. For CENELEC A band, the data rate ranges from 4.5 kbps in ROBO mode to 46 kbps with D8PSK modulation. For the FCC-above-CENELEC frequencies, throughput varies from 21 kbps in ROBO mode to a maximum of 234 kbps using D8PSK.

One major difference between IEEE 802.15.4g/e and IEEE 1901.2a is the full integration of different types of modulation and tone maps by a single PHY layer in the IEEE 1901.2a specification. IEEE 802.15.4g/e doesn't really define a multi-PHY management algorithm. The PHY payload size can change dynamically, based on channel conditions in IEEE 1901.2a. Therefore, MAC sublayer segmentation is implemented. If the size of the MAC payload is too large to fit within one PHY service data unit (PSDU), the MAC payload is partitioned into smaller segments. MAC payload segmentation is done by dividing the MAC payload into multiple smaller amounts of data (segments), based on PSDU size. The segmentation may require the addition of padding bytes to the last payload segment so that the final MPDU fills the PSDU. All forms of addressing (unicast and broadcast) are subject to segmentation.

MAC Layer

The MAC frame format of IEEE 1901.2a is based on the IEEE 802.15.4 MAC frame but integrates the latest IEEE 802.15.4e-2012 amendment, which enables key features to be supported. (For more information on the 802.15.4 MAC frame format, refer to Figure 4-6. For the 802.15.4e MAC frame format, see Figure 4-9.) One of the key components brought from 802.15.4e to IEEE 1901.2a is information elements. With IE support, additional capabilities, such as IEEE 802.15.9 Key Management Protocol and SSID, are supported. Figure 4-12 provides an overview of the general MAC frame format for IEEE 1901.2. Note that the numeric value above each field in the frame shows the size of the field, in bytes.

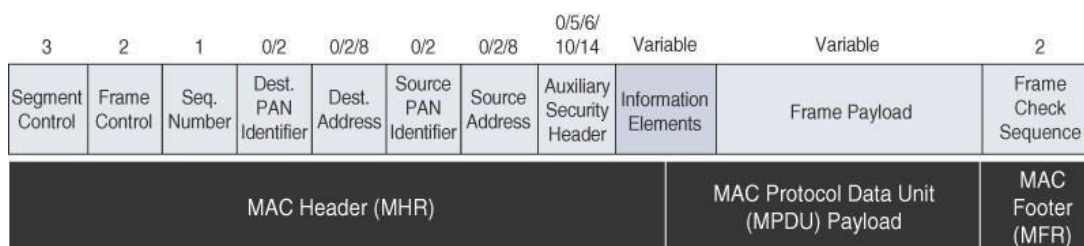


Figure 4-12 General MAC Frame Format for IEEE 1901.2

As shown in Figure 4-12, IEEE 1901.2 has a Segment Control field. This is a new field that was not present in our previous discussions of the MAC frame for 802.15.4 and 802.15.4e. This field handles

the segmentation or fragmentation of upper-layer packets with sizes larger than what can be carried in the MAC protocol data unit (MPDU).

Topology

Use cases and deployment topologies for IEEE 1901.2a are tied to the physical power lines. As with wireless technologies, signal propagation is limited by factors such as noise, interference, distortion, and attenuation. These factors become more prevalent with distance, so most NB-PLC deployments use some sort of mesh topology. Mesh networks offer the advantage of devices relaying the traffic of other devices so longer distances can be segmented. The IEEE 1901.2a standard offers the flexibility to run any upper-layer protocol. So, implementations of IPv6 6LoWPAN and RPL IPv6 protocols are supported.

Security

IEEE 1901.2a security offers similar features to IEEE 802.15.4g. Encryption and authentication are performed using AES. In addition, IEEE 1901.2a aligns with 802.15.4g in its ability to support the IEEE 802.15.9 Key Management Protocol. However, some differences exist. These differences are mostly tied to the PHY layer fragmentation capabilities of IEEE 1901.2a and include the following:

- The Security Enabled bit in the Frame Control field should be set in all MAC frames carrying segments of an encrypted frame.
- If data encryption is required, it should be done before packet segmentation. During packet encryption, the Segment Control field should not be included in the input to the encryption algorithm.
- On the receiver side, the data decryption is done after packet reassembly.
- When security is enabled, the MAC payload is composed of the ciphered payload and the message integrity code (MIC) authentication tag for non-segmented payloads. If the payload is segmented, the MIC is part of the last packet (segment) only. The MIC authentication is computed using only information from the MHR of the frame carrying the first segment.

Competitive Technologies

In the domain of NB-PLC, two technologies compete against IEEE 1901.2a: G3-PLC (now ITU G.9903) and PRIME (now ITU G.9904). Both of these technologies were initially developed to address a single use case: smart metering deployment in Europe over the CENELEC A band. IEEE 1901.2a has portions of G3-PLC and PRIME, and it also competes with them. More specifically, G3-PLC is really close to IEEE 1901.2. The main differences include the fact that G3-PLC mandates data link layer protocol options for bootstrapping and allocating device addresses, and it is incompatible with IEEE 802.15.4g/e and an end-to-end IPv6 model.

3.6.4 IEEE 802.11ah

In unconstrained networks, IEEE 802.11 Wi-Fi is certainly the most successfully deployed wireless technology. This standard is a key IoT wireless access technology, either for connecting endpoints such as fog computing nodes, high-data-rate sensors, and audio or video analytics devices or for deploying Wi-Fi backhaul infrastructures, such as outdoor Wi-Fi mesh in smart cities, oil and mining,

or other environments. However, Wi-Fi lacks sub-GHz support for better signal penetration, low power for battery-powered nodes, and the ability to support a large number of devices. For these reasons, the IEEE 802.11 working group launched a task group named IEEE 802.11ah to specify a sub-GHz version of Wi-Fi. Three main use cases are identified for IEEE 802.11ah:

- **Sensors and meters covering a smart grid:** Meter to pole, environmental/agricultural monitoring, industrial process sensors, indoor healthcare system and fitness sensors, home and building automation sensors.
- **Backhaul aggregation of industrial sensors and meter data:** Potentially connecting IEEE 802.15.4g subnetworks,
- **Extended range Wi-Fi:** For outdoor extended-range hotspot or cellular traffic offloading when distances already covered by IEEE 802.11a/b/g/n/ac are not good enough.

Standardization and Alliances

For the 802.11ah standard, the Wi-Fi Alliance defined a new brand called Wi-Fi HaLow. This marketing name is based on a play on words between “11ah” in reverse and “low power.” It is similar to the word “hello” but it is pronounced “hay-low.” The HaLow brand exclusively covers IEEE 802.11ah for sub-GHz device certification. Wi-Fi HaLow is a commercial designation for products incorporating IEEE 802.11ah technology.

Physical Layer

IEEE 802.11ah essentially provides an additional 802.11 physical layer operating in unlicensed sub-GHz bands. For example, various countries and regions use the following bands for IEEE 802.11ah: 868–868.6 MHz for EMEAR, 902–928 MHz and associated subsets for North America and Asia-Pacific regions, and 314–316 MHz, 430–434 MHz, 470–510 MHz, and 779–787 MHz for China.

Based on OFDM modulation, IEEE 802.11ah uses channels of 2, 4, 8, or 16 MHz (and also 1 MHz for low-bandwidth transmission). This is one-tenth of the IEEE 802.11ac channels, resulting in one-tenth of the corresponding data rates of IEEE 802.11ac. The IEEE 802.11ac standard is a high-speed wireless LAN protocol at the 5 GHz band that is capable of speeds up to 1 Gbps. While 802.11ah does not approach this transmission speed (as it uses one-tenth of 802.11ac channel width, it reaches one-tenth of 802.11ac speed), it does provide an extended range for its lower speed data. For example, at a data rate of 100 kbps, the outdoor transmission range for IEEE 802.11ah is expected to be 0.62 mile.

MAC Layer

The IEEE 802.11ah MAC layer is optimized to support the new sub-GHz Wi-Fi PHY while providing low power consumption and the ability to support a larger number of endpoints. Enhancements and features specified by IEEE 802.11ah for the MAC layer include the following:

- **Number of devices:** Has been scaled up to 8192 per access point.
- **MAC header:** Has been shortened to allow more efficient communication.
- **Null data packet (NDP) support:** Is extended to cover several control and management frames. Relevant information is concentrated in the PHY header and the additional overhead associated with decoding the MAC header and data payload is avoided. This change makes the control frame exchanges efficient and less power-consuming for the receiving stations.
- **Grouping and sectorization:** Enables an AP to use sector antennas and also group stations (distributing a group ID). In combination with RAW and TWT, this mechanism reduces

contention in large cells with many clients by restricting which group, in which sector, can contend during which time window. (Sectors are described in more detail in the following section.)

- **Restricted access window (RAW):** Is a control algorithm that avoids simultaneous transmissions when many devices are present and provides fair access to the wireless network. By providing more efficient access to the medium, additional power savings for battery-powered devices can be achieved, and collisions are reduced.
- **Target wake time (TWT):** Reduces energy consumption by permitting an access point to define times when a device can access the network. This allows devices to enter a low-power state until their TWT time arrives. It also reduces the probability of collisions in large cells with many clients.
- **Speed frame exchange:** Enables an AP and endpoint to exchange frames during a reserved transmit opportunity (TXOP). This reduces contention on the medium, minimizes the number of frame exchanges to improve channel efficiency, and extends battery life by keeping awake times short.

From the above feature list the 802.11ah MAC layer is focused on power consumption and mechanisms to allow low-power Wi-Fi stations to wake up less often and operate more efficiently. This sort of MAC layer is ideal for IoT devices that often produce short, low-bit-rate transmissions.

Topology

While IEEE 802.11ah is deployed as a star topology, it includes a simple hops relay operation to extend its range. This relay option is not capped, but the IEEE 802.11ah task group worked on the assumption of two hops. It allows one 802.11ah device to act as an intermediary and relay data to another. In some ways, this is similar to a mesh, and it is important to note that the clients and not the access point handle the relay function. This relay operation can be combined with a higher transmission rate or modulation and coding scheme (MCS). This means that a higher transmit rate is used by relay devices talking directly to the access point.

Sectorization is a technique that involves partitioning the coverage area into several sectors to get reduced contention within a certain sector. This technique is useful for limiting collisions in cells that have many clients. This technique is also often necessary when the coverage area of 802.11ah access points is large, and interference from neighboring access points is problematic. Sectorization uses an antenna array and beam-forming techniques to partition the cell-coverage area. Figure 4-14 shows an example of 802.11ah sectorization.

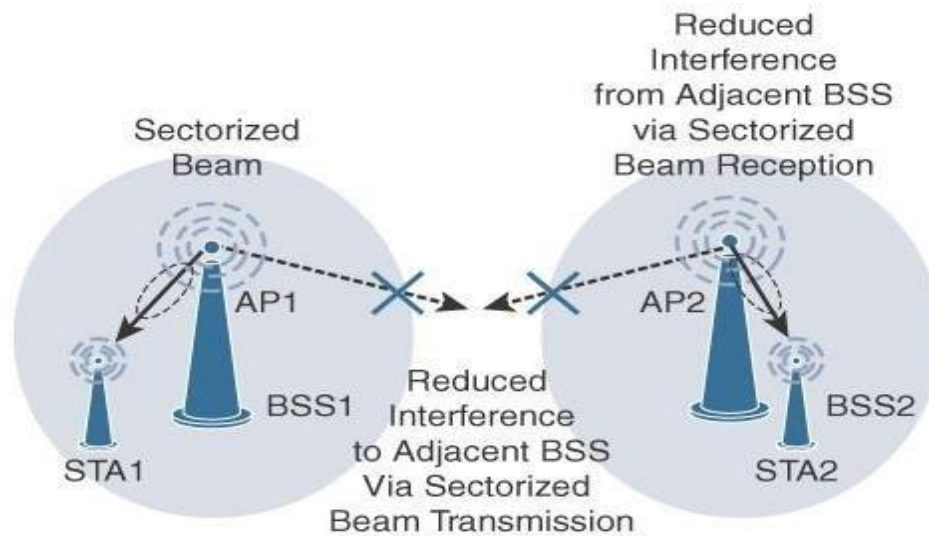


Figure 4-14 IEEE 802.11ah Sectorization

Security

No additional security has been identified for IEEE 802.11ah compared to other IEEE 802.11 specifications. These protocols include IEEE 802.15.4, IEEE 802.15.4e, and IEEE 1901.2a, and the security information for them is also applicable to IEEE 802.11ah.

Competitive Technologies

Competitive technologies to IEEE 802.11ah are IEEE 802.15.4 and IEEE 802.15.4e, along with the competitive technologies highlighted in each of their sections.

3.6.5 LoRaWAN

In recent years, a new set of wireless technologies known as Low-Power Wide-Area (LPWA) has received a lot of attention from the industry and press. Particularly well adapted for long-range and battery-powered endpoints, LPWA technologies open new business opportunities to both services providers and enterprises considering IoT solutions. This section discusses an example of an unlicensed-band LPWA technology, known as LoRaWAN, and the next section, “NB-IoT and Other LTE Variations,” reviews licensed-band alternatives from the 3rd Generation Partnership Project (3GPP).

Standardization and Alliances

Initially, LoRa was a physical layer, or Layer 1, modulation that was developed by a French company named Cycleo. Later, Cycleo was acquired by Semtech. Optimized for long-range, two-way communications and low power consumption, the technology evolved from Layer 1 to a broader scope through the creation of the LoRa Alliance. It quickly achieved industry support and currently has hundreds of members.

Semtech LoRa as a Layer 1 PHY modulation technology is available through multiple chipset vendors. To differentiate from the physical layer modulation known as LoRa, the LoRa Alliance uses

the term LoRaWAN to refer to its architecture and its specifications that describe end-to-end LoRaWAN communications and protocols. Figure 4-15 provides a high-level overview of the LoRaWAN layers. In this figure, notice that Semtech is responsible for the PHY layer, while the LoRa Alliance handles the MAC layer and regional frequency bands.

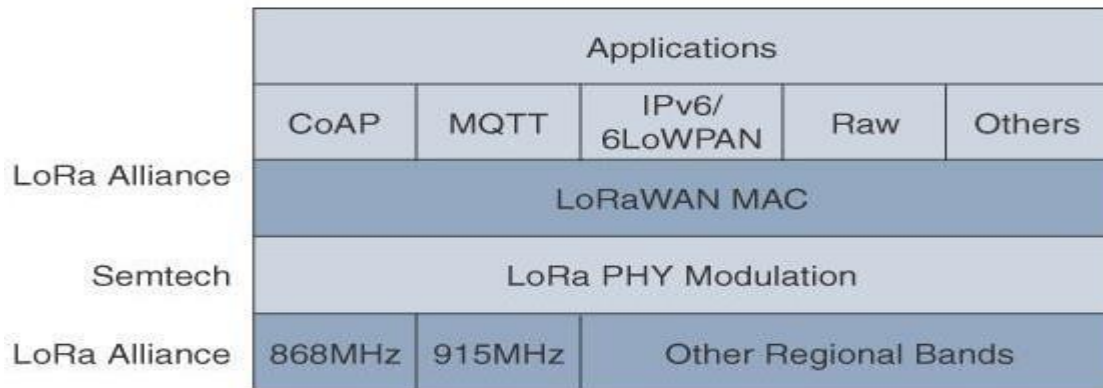


Figure 4-15 *LoRaWAN Layers*

Overall, the LoRa Alliance owns and manages the roadmap and technical development of the LoRaWAN architecture and protocol. This alliance also handles the LoRaWAN endpoint certification program and technology promotion through its certification and marketing committees.

Physical Layer

Semtech LoRa modulation is based on chirp spread spectrum modulation, which trades a lower data rate for receiver sensitivity to significantly increase the communication distance. In addition, it allows demodulation below the noise floor, offers robustness to noise and interference, and manages a single channel occupation by different spreading factors. This enables LoRa devices to receive on multiple channels in parallel.

LoRaWAN 1.0.2 regional specifications describe the use of the main unlicensed sub-GHz frequency bands of 433 MHz, 779–787 MHz, 863–870 MHz, and 902–928 MHz, as well as regional profiles for a subset of the 902–928 MHz bandwidth. For example, Australia utilizes 915–928 MHz frequency bands, while South Korea uses 920–923 MHz and Japan uses 920–928 MHz.

The data rate in LoRaWAN varies depending on the frequency bands and adaptive data rate (ADR). ADR is an algorithm that manages the data rate and radio signal for each endpoint. The ADR algorithm ensures that packets are delivered at the best data rate possible and that network performance is both optimal and scalable. Endpoints close to the gateways with good signal values transmit with the highest data rate, which enables a shorter transmission time over the wireless network, and the lowest transmit power. Meanwhile, endpoints at the edge of the link budget communicate at the lowest data rate and highest transmit power.

An important feature of LoRa is its ability to handle various data rates via the spreading factor. Devices with a low spreading factor (SF) achieve less distance in their communications but transmit at faster speeds, resulting in less airtime. A higher SF provides slower transmission rates but achieves a higher reliability at longer distances. Table 4-4 illustrates how LoRaWAN data rates can vary depending on the associated spreading factor for the two main frequency bands, 863–870 MHz and 902–928 MHz.

Configuration	863–870 MHz bps	902–928 MHz bps
LoRa: SF12/125 kHz	250	N/A
LoRa: SF11/125 kHz	440	N/A
LoRa: SF10/125 kHz	980	980
LoRa: SF9/125 kHz	1760	1760
LoRa: SF8/125 kHz	3125	3125
LoRa: SF7/125 kHz	5470	5470
LoRa: SF7/250 kHz	11,000	N/A
FSK: 50 kbps	50,000	N/A
LoRa: SF12/500 kHz	N/A	980
LoRa: SF11/500 kHz	N/A	1760
LoRa: SF10/500 kHz	N/A	3900
LoRa: SF9/500 kHz	N/A	7000
LoRa: SF8/500 kHz	N/A	12,500
LoRa: SF7/500 kHz	N/A	21,900

Table 4-4 LoRaWAN Data Rate Example

In Table 4-4, notice the relationship between SF and data rate. For example, at an SF value of 12 for 125 kHz of channel bandwidth, the data rate is 250 bps. However, when the SF is decreased to a value of 7, the data rate increases to 5470 bps. Channel bandwidth values of 125 kHz, 250 kHz, and 500 kHz are also evident in Table 4-4. The effect of increasing the bandwidth is that faster data rates can be achieved for the same spreading factor.

MAC Layer

As mentioned previously, the MAC layer is defined in the LoRaWAN specification. This layer takes advantage of the LoRa physical layer and classifies LoRaWAN endpoints to optimize their battery life and ensure downstream communications to the LoRaWAN endpoints. The LoRaWAN specification documents three classes of LoRaWAN devices:

- **Class A:** This class is the default implementation. Optimized for battery-powered nodes, it allows bidirectional communications, where a given node is able to receive downstream traffic after transmitting. Two receive windows are available after each transmission.
- **Class B:** This class was designated “experimental” in LoRaWAN 1.0.1 until it can be better defined. A Class B node or endpoint should get additional receive windows compared to Class A, but gateways must be synchronized through a beaconing process.
- **Class C:** This class is particularly adapted for powered nodes. This classification enables a node to be continuously listening by keeping its receive window open when not transmitting.

LoRaWAN messages, either uplink or downlink, have a PHY payload composed of a 1-byte MAC header, a variable-byte MAC payload, and a MIC that is 4 bytes in length. The MAC payload size depends on the frequency band and the data rate, ranging from 59 to 230 bytes for the 863–870 MHz band and 19 to 250 bytes for the 902–928 MHz band. Figure 4-16 shows a high-level LoRaWAN MAC frame format.



Figure 4-16 *High-Level LoRaWAN MAC Frame Format*

In version 1.0.x, LoRaWAN utilizes six MAC message types. LoRaWAN devices use join request and join accept messages for over-the-air (OTA) activation and joining the network. The other message types are unconfirmed data up/down and confirmed data up/down. A “confirmed” message is one that must be acknowledged, and “unconfirmed” signifies that the end device does not need to acknowledge. “up/down” is simply a directional notation identifying whether the message flows in the uplink or downlink path. Uplink messages are sent from endpoints to the network server and are relayed by one or more LoRaWAN gateways. Downlink messages flow from the network server to a single endpoint and are relayed by only a single gateway. Multicast over LoRaWAN is being considered for future versions.

LoRaWAN endpoints are uniquely addressable through a variety of methods, including the following:

- An endpoint can have a global end device ID or DevEUI represented as an IEEE EUI-64 address.
- An endpoint can have a global application ID or AppEUI represented as an IEEE EUI-64 address that uniquely identifies the application provider, such as the owner, of the end device.
- In a LoRaWAN network, endpoints are also known by their end device address, known as a DevAddr, a 32-bit address. The 7 most significant bits are the network identifier (NwkID), which identifies the LoRaWAN network. The 25 least significant bits are used as the network address (NwkAddr) to identify the endpoint in the network.

Topology

LoRaWAN topology is often described as a “star of stars” topology. As shown in Figure 4-17, the infrastructure consists of endpoints exchanging packets through gateways acting as bridges, with a central LoRaWAN network server. Gateways connect to the backend network using standard IP connections, and endpoints communicate directly with one or more gateways.

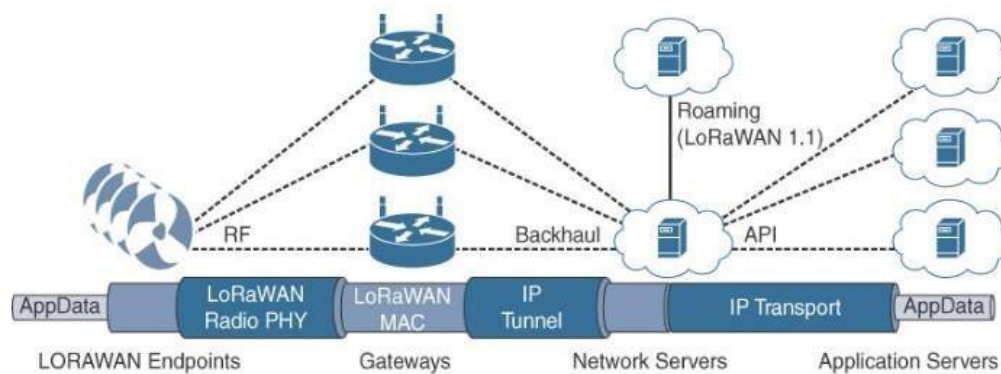


Figure 4-17 LoRaWAN Architecture

In Figure 4-17, LoRaWAN endpoints transport their selected application data over the LoRaWAN MAC layer on top of one of the supported PHY layer frequency bands. The application data is contained in upper protocol layers. These upper layers are not the responsibility of the LoRa Alliance, but best practices may be developed and recommended. These upper layers could just be raw data on top of the LoRaWAN MAC layer, or the data could be stacked in multiple protocols. Figure 4-17 also shows how LoRaWAN gateways act as bridges that relay between endpoints and the network servers. Multiple gateways can receive and transport the same packets. When duplicate packets are received, de-duplication is a function of the network server.

The LoRaWAN network server manages the data rate and radio frequency (RF) of each endpoint through the adaptive data rate (ADR) algorithm. ADR is a key component of the network scalability, performance, and battery life of the endpoints. The LoRaWAN network server forwards application data to the application servers, as depicted in Figure 4-17.

In future versions of the LoRaWAN specification, roaming capabilities between LoRaWAN network servers will be added. These capabilities will enable mobile endpoints to connect and roam between different LoRaWAN network infrastructures.

Security

Security in a LoRaWAN deployment applies to different components of the architecture, as detailed in Figure 4-18. LoRaWAN endpoints must implement two layers of security, protecting communications and data privacy across the network.

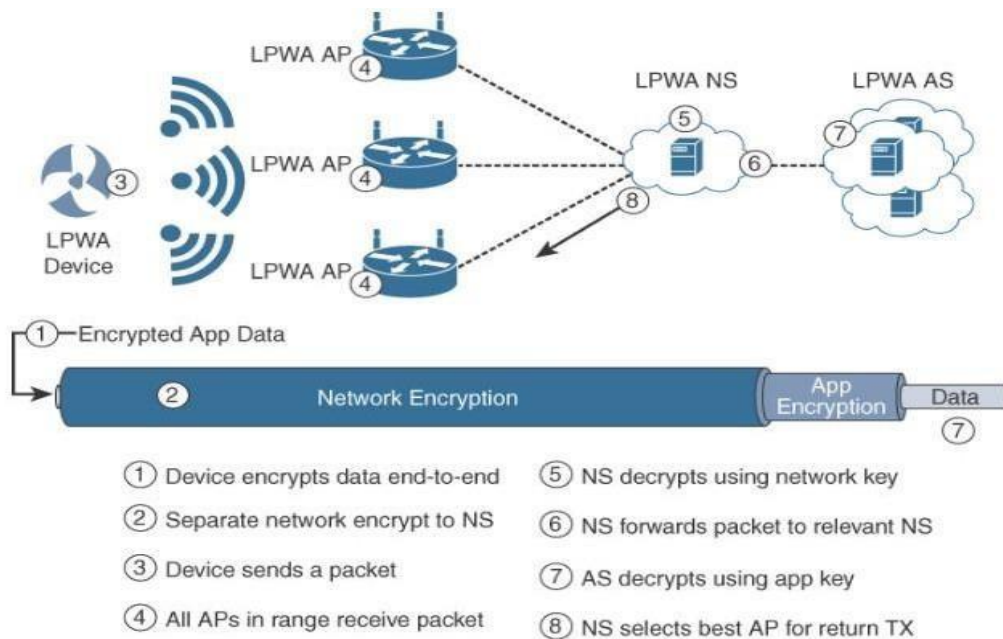


Figure 4-18 LoRaWAN Security

The first layer, called “network security” but applied at the MAC layer, guarantees the authentication of the endpoints by the LoRaWAN network server. Also, it protects LoRaWAN packets by performing encryption based on AES. Each endpoint implements a network session key (NwkSKey), used by both itself and the LoRaWAN network server. The NwkSKey ensures data integrity through computing and checking the MIC of every data message as well as encrypting and decrypting MAC-only data message payloads.

The second layer is an application session key (AppSKey), which performs encryption and decryption functions between the endpoint and its application server. Furthermore, it computes and checks the application-level MIC, if included. This ensures that the LoRaWAN service provider does not have access to the application payload if it is not allowed that access.

Endpoints receive their AES-128 application key (AppKey) from the application owner. This key is most likely derived from an application-specific root key exclusively known to and under the control of the application provider. For production deployments, it is expected that the LoRaWAN gateways are protected as well, for both the LoRaWAN traffic and the network management and operations over their backhaul link(s). This can be done using traditional VPN and IPsec technologies that demonstrate scaling in traditional IT deployments.

LoRaWAN endpoints attached to a LoRaWAN network must get registered and authenticated. This can be achieved through one of the two join mechanisms:

- **Activation by personalization (ABP):** Endpoints don’t need to run a join procedure as their individual details, including DevAddr and the NwkSKey and AppSKey session keys, are

preconfigured and stored in the end device. This same information is registered in the LoRaWAN network server.

- **Over-the-air activation (OTAA):** Endpoints are allowed to dynamically join a particular LoRaWAN network after successfully going through a join procedure. The join procedure must be done every time a session context is renewed. During the join process, which involves the sending and receiving of MAC layer join request and join accept messages, the node establishes its credentials with a LoRaWAN network server, exchanging its globally unique DevEUI, AppEUI, and AppKey. The AppKey is then used to derive the session NwkSKey and AppSKey keys.

Competitive Technologies

LPWA solutions and technologies are split between unlicensed and licensed bands. The licensed- band technologies are dedicated to mobile service providers that have acquired spectrum licenses; they are discussed in the next section. In addition, several technologies are targeting the unlicensed- band LPWA market to compete against LoRaWAN. The LPWA market is quickly evolving. Table 4- 5 evaluates two of the best-established vendors known to provide LPWA options.

Characteristic	LoRaWAN	Sigfox	Ingenu Onramp
Frequency bands	433 MHz, 868 MHz, 902–928 MHz	433 MHz, 868 MHz, 902–928 MHz	2.4 GHz
Modulation	Chirp spread spectrum	Ultra-narrowband	DSSS
Topology	Star of stars	Star	Star; tree supported with an RPMA extender
Data rate	250 bps–50 kbps (868 MHz) 980 bps–21.9 kbps (915 MHz)	100 bps (868 MHz) 600 bps (915 MHz)	6 kbps
Adaptive data rate	Yes	No	No
Payload	59–230 bytes (868 MHz) 19–250 bytes (915 MHz)	12 bytes	6 bytes–10 KB
Two-way communications	Yes	Partial	Yes
Geolocation	Yes (LoRa GW version 2 reference design)	No	No
Roaming	Yes (LoRaWAN 1.1)	No	Yes
Specifications	LoRA Alliance	Proprietary	Proprietary

Table 4-5 Unlicensed LPWA Technology Comparison

3.6.6 NB-IoT and Other LTE Variations

Existing cellular technologies, such as GPRS, Edge, 3G, and 4G/LTE, are not particularly well adapted to battery-powered devices and small objects specifically developed for the Internet of Things. While industry players have been developing unlicensed-band LPWA technologies, 3GPP and associated vendors have been working on evolving cellular technologies to better address IoT requirements. The effort started with the definition of new LTE device categories. The aim was to both align with specific IoT requirements, such as low throughput and low power consumption, and decrease the complexity and cost of the LTE devices. This resulted in the definition of the LTE-M work item.

Because the new LTE-M device category was not sufficiently close to LPWA capabilities, in 2015 3GPP approved a proposal to standardize a new narrowband radio access technology called Narrowband IoT (NB-IoT). NB-IoT specifically addresses the requirements of a massive number of low-throughput devices, low device power consumption, improved indoor coverage, and optimized network architecture. The following sections review the proposed evolution of cellular technologies to better support the IoT opportunities by mobile service providers.

Standardization and Alliances

The 3GPP organization includes multiple working groups focused on many different aspects of telecommunications (for example, radio, core, terminal, and so on). Many service providers and vendors make up 3GPP, and the results of their collaborative work in these areas are the 3GPP specifications and studies. The workflow within 3GPP involves receiving contributions related to licensed LPWA work from the involved vendors. Then, depending on the access technology that is most closely aligned, such as 3G, LTE, or GSM, the IoT-related contribution is handled by either 3GPP or the GSM EDGE Radio Access Networks (GERAN) group.

Mobile vendors and service providers are not willing to lose leadership in this market of connecting IoT devices. Therefore, a couple intermediate steps have been pushed forward, leading to the final objectives set for NB-IoT and documented by 3GPP. At the same time, another industry group, the GSM Association (GSMA), has proposed the Mobile IoT Initiative, which “is designed to accelerate the commercial availability of LPWA solutions in licensed spectrum.” For more information on the Mobile IoT Initiative, go to www.gsma.com/connectedliving/mobile-iot-initiative/.

3.6.6.1 LTE Cat 0

The first enhancements to better support IoT devices in 3GPP occurred in LTE Release 12. A new user equipment (UE) category, Category 0, was added, with devices running at a maximum data rate of 1 Mbps. Generally, LTE enhancements target higher bandwidth improvements. Category 0 includes important characteristics to be supported by both the network and end devices. Meanwhile, the UE still can operate in existing LTE systems with bandwidths up to 20 MHz. These Cat 0 characteristics include the following:

- **Power saving mode (PSM):** This new device status minimizes energy consumption. Energy consumption is expected to be lower with PSM than with existing idle mode. PSM is defined as being similar to “powered off” mode, but the device stays registered with the network. By staying registered, the device avoids having to reattach or reestablish its network connection. The device negotiates with the network the idle time after which it will wake up. When it wakes up, it initiates a tracking area update (TAU), after which it stays available for a configured time and then switches back to sleep mode or PSM. A TAU is a procedure that an LTE device uses to let the network know its current tracking area, or the group of towers in the network from which it can be reached. Basically, with PSM, a device can be practically powered off but not lose its place in the network.
- **Half-duplex mode:** This mode reduces the cost and complexity of a device’s implementation because a duplex filter is not needed. Most IoT endpoints are sensors that send low amounts of data that do not have a full-duplex communication requirement.

3.6.6.2 LTE-M

Following LTE Cat 0, the next step in making the licensed spectrum more supportive of IoT devices was the introduction of the LTE-M category for 3GPP LTE Release 13. These are the main characteristics of the LTE-M category in Release 13:

- **Lower receiver bandwidth:** Bandwidth has been lowered to 1.4 MHz versus the usual 20 MHz. This further simplifies the LTE endpoint.
- **Lower data rate:** Data is around 200 kbps for LTE-M, compared to 1 Mbps for Cat 0.
- **Half-duplex mode:** Just as with Cat 0, LTE-M offers a half-duplex mode that decreases node complexity and cost.
- **Image Enhanced discontinuous reception (eDRX):** This capability increases from seconds to minutes the amount of time an endpoint can “sleep” between paging cycles. A paging cycle is a periodic check-in with the network. This extended “sleep” time between paging cycles extends the battery lifetime for an endpoint significantly.

LTE-M requires new chipsets and additional software development. Commercial deployment is expected in 2017. Mobile carriers expect that only new LTE-M software will be required on the base stations, which will prevent re-investment in hardware.

3.6.6.3 NB-IoT

Recognizing that the definition of new LTE device categories was not sufficient to support LPWA IoT requirement, 3GPP specified Narrowband IoT (NB-IoT). The work on NB-IoT started with multiple proposals pushed by the involved vendors, including the following:

- Extended Coverage GSM (EC-GSM), Ericsson proposal
- Narrowband GSM (N-GSM), Nokia proposal
- Narrowband M2M (NB-M2M), Huawei/Neul proposal
- Narrowband OFDMA (orthogonal frequency-division multiple access), Qualcomm proposal
- Narrowband Cellular IoT (NB-CIoT), combined proposal of NB-M2M and NB-OFDMA
- Narrowband LTE (NB-LTE), Alcatel-Lucent, Ericsson, and Nokia proposal

- Cooperative Ultra Narrowband (C-UNB), Sigfox proposal

Consolidation occurred with the agreement to specify a single NB-IoT version based on orthogonal frequency-division multiple access (OFDMA) in the downlink and a couple options for the uplink. OFDMA is a modulation scheme in which individual users are assigned subsets of subcarrier frequencies. This enables multiple users to transmit low-speed data simultaneously. For more information on the uplink options, refer to the 3GPP specification TR 36.802.

Three modes of operation are applicable to NB-IoT:

- **Standalone:** A GSM carrier is used as an NB-IoT carrier, enabling reuse of 900 MHz or 1800 MHz.
- **In-band:** Part of an LTE carrier frequency band is allocated for use as an NB-IoT frequency. The service provider typically makes this allocation, and IoT devices are configured accordingly. Be aware that if these devices must be deployed across different countries or regions using a different service provider, problems may occur unless there is some coordination between the service providers, and the NB-IoT frequency band allocations are the same.
- **Guard band:** An NB-IoT carrier is between the LTE or WCDMA bands. This requires coexistence between LTE and NB-IoT bands.

In its Release 13, 3GPP completed the standardization of NB-IoT. Beyond the radio-specific aspects, this work specifies the adaptation of the IoT core to support specific IoT capabilities, including simplifying the LTE attach procedure so that a dedicated bearer channel is not required and transporting non-IP data. Subsequent releases of 3GPP NB-IoT will introduce additional features and functionality, such as multicasting, and will be backward compatible with Release 13.

Mobile service providers consider NB-IoT the target technology as it allows them to leverage their licensed spectrum to support LPWA use cases. For instance, NB-IoT is defined for a 200-kHz-wide channel in both uplink and downlink, allowing mobile service providers to optimize their spectrum, with a number of deployment options for GSM, WCDMA, and LTE spectrum, as shown in Figure 4-19.

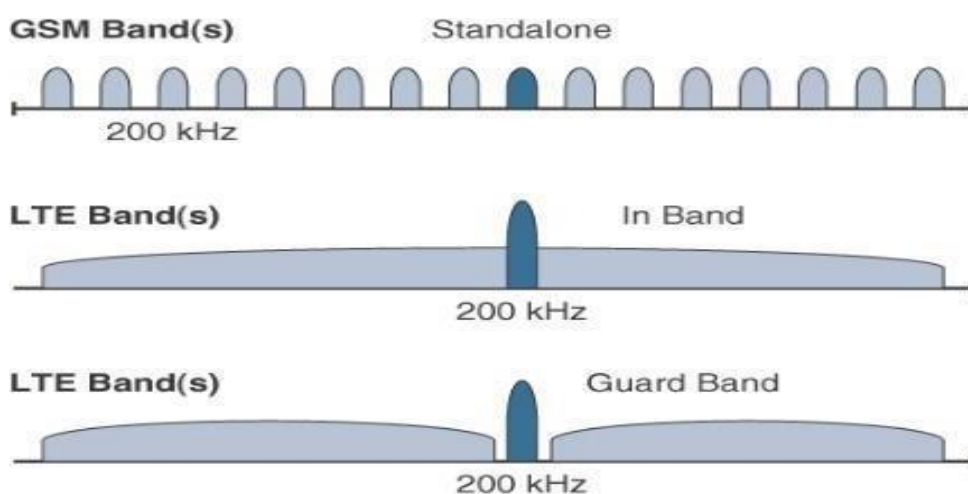


Figure 4-19 NB-IoT Deployment Options

In an LTE network, resource blocks are defined with an effective bandwidth of 180 kHz, while on NB-IoT, tone or subcarriers replace the LTE resource blocks. The uplink channel can be 15 kHz or

3.75 kHz or multi-tone ($n \times 15$ kHz, n up to 12). At Layer 1, the maximum transport block size (TBS) for downlink is 680 bits, while uplink is 1000 bits. At Layer 2, the maximum Packet Data Convergence Protocol (PDCP) service data unit (SDU) size is 1600 bytes. NB-IoT operates in half-duplex frequency-division duplexing (FDD) mode with a maximum data rate uplink of 60 kbps and downlink of 30 kbps.

Topology

NB-IoT is defined with a link budget of 164 dB; compare this with the GPRS link budget of 144 dB, used by many machine-to-machine services. The additional 20 dB link budget increase should guarantee better signal penetration in buildings and basements while achieving battery life requirements.

Competitive Technologies

In licensed bands, it is expected that 3GPP NB-IoT will be the adopted LPWA technology when it is fully available. Competitive technologies are mostly the unlicensed-band LPWA technologies such as LoRaWAN. The main challenge faced by providers of the licensed bands is the opportunity for non-mobile service providers to grab market share by offering IoT infrastructure without buying expensive spectrum.