Introduction to Wireless Networks

PERFORMANCE OF WIRELESS NETWORKS, CHAPTER 5

Ubiquitous Connectivity §

One of the most transformative technology trends of the past decade is the availability and growing expectation of ubiquitous connectivity. Whether it is for checking email, carrying a voice conversation, web browsing, or myriad other use cases, we now expect to be able to access these online services regardless of location, time, or circumstance: on the run, while standing in line, at the office, on a subway, while in flight, and everywhere in between. Today, we are still often forced to be proactive about finding connectivity (e.g., looking for a nearby WiFi hotspot) but without a doubt, the future is about ubiquitous connectivity where access to the Internet is omnipresent.

Wireless networks are at the epicenter of this trend. At its broadest, a wireless network refers to any network not connected by cables, which is what enables the desired convenience and mobility for the user. Not surprisingly, given the myriad different use cases and applications, we should also expect to see dozens of different wireless technologies to meet the needs, each with its own performance characteristics and each optimized for a specific task and context. Today, we already have over a dozen widespread wireless technologies in use: WiFi, Bluetooth, ZigBee, NFC, WiMAX, LTE, HSPA, EV-DO, earlier 3G standards, satellite services, and more.

As such, given the diversity, it is not wise to make sweeping generalizations about performance of wireless networks. However, the good news is that most wireless technologies operate on common principles, have common trade-offs, and are subject to common performance criteria and constraints. Once we uncover and understand these fundamental principles of wireless performance, most of the other pieces will begin to automatically fall into place.

Further, while the mechanics of data delivery via radio communication are fundamentally different from the tethered world, the outcome as experienced by the user is, or should be, all the same—same performance, same results. In the long run all applications are and will be delivered over wireless networks; it just may be the case that some will be accessed more frequently over wireless than others. There is no such thing as a *wired application*, and there is zero demand for such a distinction.

All applications should perform well regardless of underlying connectivity. As a user, you should not care about the underlying technology in use, but as developers we must think ahead and architect our applications to anticipate the differences between the different types of networks. And the good news is every optimization that we apply for wireless networks will translate to a better experience in all other contexts. Let’s dive in.

Types of Wireless Networks

A network is a group of devices connected to one another. In the case of wireless networks, radio communication is usually the medium of choice. However, even within the radio-powered subset, there are dozens of different technologies designed for use at different scales, topologies, and for dramatically different use cases. One way to illustrate this difference is to partition the use cases based on their "geographic range":

Type	Range	Applications	Standards
Personal area network (PAN)	Within reach of a person	Cable replacement for peripherals	Bluetooth, ZigBee, NFC
Local area network (LAN)	Within a building or campus	Wireless extension of wired network	IEEE 802.11 (WiFi)
Metropolitan area network (MAN)	Within a city	Wireless inter-network connectivity	IEEE 802.15 (WiMAX)
Wide area network (WAN)	Worldwide	Wireless network access	Cellular (UMTS, LTE, etc.)

Table 5-1. Types of wireless networks

The preceding classification is neither complete nor entirely accurate. Many technologies and standards start within a specific use case, such as Bluetooth for PAN applications and cable replacement, and with time acquire more capabilities, reach, and throughput. In fact, the latest drafts of Bluetooth now provide seamless interoperability with 802.11 (WiFi) for high-bandwidth use cases. Similarly, technologies such as WiMAX have their origins as fixed-wireless solutions, but with time acquired additional mobility capabilities, making them a viable alternative to other WAN and cellular technologies.

The point of the classification is not to partition each technology into a separate bin, but to highlight the high-level differences within each use case. Some devices have access to a continuous power source; others must optimize their battery life at all costs. Some require Gbit/s+ data rates; others are built to transfer tens or hundreds of bytes of data (e.g., NFC). Some applications require always-on connectivity, while others are delay and latency tolerant. These and a large number of other criteria are what determine the original characteristics of each type of network. However, once in place, each standard continues to evolve: better battery capacities, faster processors, improved modulation algorithms, and other advancements continue to extend the use cases and performance of each wireless standard.

Note

Your next application may be delivered over a mobile network, but it may also rely on NFC for payments, Bluetooth for P2P communication via WebRTC, and WiFi for HD streaming. It is not

Performance Fundamentals of Wireless Networks §

Each and every type of wireless technology has its own set of constraints and limitations. However, regardless of the specific wireless technology in use, all communication methods have a maximum channel capacity, which is determined by the same underlying principles. In fact, Claude E. Shannon gave us an exact mathematical model (Channel capacity is the maximum information rate) to determine channel capacity, regardless of the technology in use.

Channel capacity is the maximum information rate

$$C = BW \times \log_2\left(1 + \frac{S}{N}\right)$$

- `C` is the channel capacity and is measured in bits per second.
- `BW` is the available bandwidth, and is measured in hertz.
- `S` is signal and `N` is noise, and they are measured in watts.

Although somewhat simplified, the previous formula captures all the essential insights we need to understand the performance of most wireless networks. Regardless of the name, acronym, or the revision number of the specification, the two fundamental constraints on achievable data rates are the amount of available bandwidth and the signal power between the receiver and the sender.

Bandwidth §

Unlike the tethered world, where a dedicated wire can be run between each network peer, radio communication by its very nature uses a shared medium: radio waves, or if you prefer, electromagnetic radiation. Both the sender and receiver must agree up-front on the specific frequency range over which the communication will occur; a well-defined range allows seamless interoperability between devices. For example, the 802.11b and 802.11g standards both use the 2.4–2.5 GHz band across all WiFi devices.

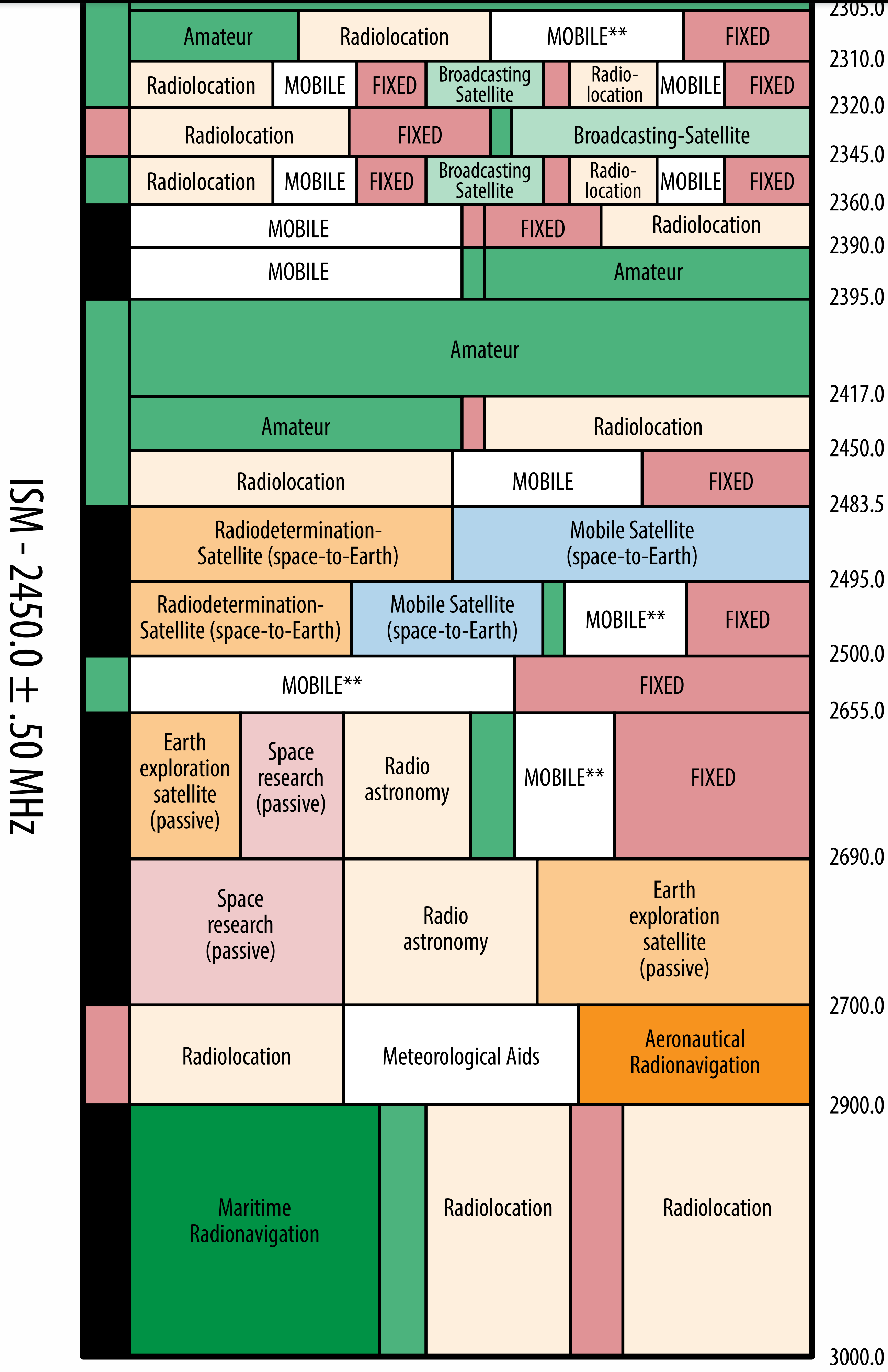
Who determines the frequency range and its allocation? In short, local government (Figure 5-1). In the United States, this process is governed by the Federal Communications Commission (FCC). In fact, due to different government regulations, some wireless technologies may work in one part of the world, but not in others. Different countries may, and often do, assign different spectrum ranges to the same wireless technology.

performance factor is the size of the assigned frequency range. As Shannon's model shows, the overall channel bitrate is directly proportional to the assigned range. Hence, all else being equal, a doubling in available frequency range will double the data rate—e.g., going from 20 to 40 MHz of bandwidth can double the channel data rate, which is exactly how 802.11n is improving its performance over earlier WiFi standards!

Finally, it is also worth noting that not all frequency ranges offer the same performance. Low-frequency signals travel farther and cover large areas (macrocells), but at the cost of requiring larger antennas and having more clients competing for access. On the other hand, high-frequency signals can transfer more data but won't travel as far, resulting in smaller coverage areas (microcells) and a requirement for more infrastructure.

Note

Certain frequency ranges are more valuable than others for some applications. Broadcast-only applications (e.g., broadcast radio) are well suited for low-frequency ranges. On the other hand, two-way communication benefits from use of smaller cells, which provide higher bandwidth and less competition.


2305.0
Amateur
Radiolocation
MOBILE**
FIXED
2310.0
Radiolocation
MOBILE
FIXED
Broadcasting Satellite
Radio-location
MOBILE
FIXED
2320.0
Radiolocation
FIXED
Broadcasting-Satellite
2345.0
Radiolocation
MOBILE
FIXED
Broadcasting Satellite
Radio-location
MOBILE
FIXED
2360.0
MOBILE
FIXED
Radiolocation
2390.0
MOBILE
Amateur
2395.0
Amateur
2417.0
Amateur
Radiolocation
2450.0
Radiolocation
MOBILE
FIXED
2483.5
ISM - 2450.0 ± .50 MHz
Radiodetermination-Satellite (space-to-Earth)
Mobile Satellite (space-to-Earth)
2495.0
Radiodetermination-Satellite (space-to-Earth)
Mobile Satellite (space-to-Earth)
MOBILE**
FIXED
2500.0
MOBILE**
FIXED
2655.0
Earth exploration satellite (passive)
Space research (passive)
Radio astronomy
MOBILE**
FIXED
2690.0
Space research (passive)
Radio astronomy
Earth exploration satellite (passive)
2700.0
Radiolocation
Meteorological Aids
Aeronautical Radionavigation
2900.0
Maritime Radionavigation
Radiolocation
Radiolocation
3000.0

A Brief History of Worldwide Spectrum Allocation and Regulation §

If you spend any time in the world of wireless communication, you will inevitably stumble into numerous debates on the state and merits of current spectrum allocation and regulation processes. But what is the history?

In the early days of radio, anyone could use any frequency range for whatever purpose she desired. All of that changed when the Radio Act of 1912 was signed into law within the United States and mandated licensed use of the radio spectrum. The original bill was in part motivated by the investigation into the sinking of the Titanic. Some speculate that the disaster could have been averted, or more lives could have been saved, if proper frequencies were monitored by all nearby vessels. Regardless, this new law set a precedent for international and federal legislation of wireless communication. Other countries followed.

A few decades later, the Communications Act of 1934 created the Federal Communications Commission (FCC), and the FCC has been responsible for managing the spectrum allocation within the U.S ever since, effectively "zoning" it by subdividing into ever-smaller parcels designed for exclusive use.

A good example of the different allocations are the "industrial, scientific, and medical" (ISM) radio bands, which were first established at the International Telecommunications Conference in 1947, and as the name implies, were reserved internationally. Both the 2.4–2.5 GHz (100 MHz) and 5.725–5.875 GHz (150 MHz) bands, which power much of our modern wireless communication (e.g., WiFi) are part of the ISM band. Further, both of these ISM bands are also considered "unlicensed spectrum," which allow anyone to operate a wireless network—for commercial or private use—in these bands as long as the hardware used respects specified technical requirements (e.g., transmit power).

Finally, due to the rising demand in wireless communication, many governments have begun to hold "spectrum auctions," where a license is sold to transmit signals over the specific bands. While examples abound, the 700 MHz FCC auction, which took place in 2008, is a good illustration: the 698–806 MHz range within the U.S. was auctioned off for a total of $19.592 billion to over a dozen different bidders (the range was subdivided into blocks). Yes, that is billion with a "b."

Bandwidth is a scarce and expensive commodity. Whether the current allocation process is fair is a subject on which much ink has been spilled and many books have been published. Looking forward, there is one thing we can be sure of: it will continue to be a highly contested area of discussion.

Signal Power §

Besides bandwidth, the second fundamental limiting factor in all wireless communication is the signal power between the sender and receiver, also known as the signal-power-to-noise-power, S/N ratio, or SNR. In essence, it is a measure that compares the level of desired signal to the level the signal has to be to carry the information.

By its very nature, all radio communication is done over a shared medium, which means that other devices may generate unwanted interference. For example, a microwave oven operating at 2.5 GHz may overlap with the frequency range used by WiFi, creating cross-standard interference. However, other WiFi devices, such as your neighbors' WiFi access point, and even your coworker's laptop accessing the same WiFi network, also create interference for your transmissions.

In the ideal case, you would be the one and only user within a certain frequency range, with no other background noise or interference. Unfortunately, that's unlikely. First, bandwidth is scarce, and second, there are simply too many wireless devices to make that work. Instead, to achieve the desired data rate where interference is present, we can either increase the transmit power, thereby increasing the strength of the signal, or decrease the distance between the transmitter and the receiver—or both, of course.

Note

Path loss, or path attenuation, is the reduction in signal power with respect to distance traveled—the exact reduction rate depends on the environment. A full discussion on this is outside the scope of this book, but if you are curious, consult your favorite search engine.

To illustrate the relationship between signal, noise, transmit power, and distance, imagine you are in a small room and talking to someone 20 feet away. If nobody else is present, you can hold a conversation at normal volume. However, now add a few dozen people into the same room, such as at a crowded party, each carrying their own conversations. All of the sudden, it would be impossible for you to hear your peer! Of course, you could start speaking louder, but doing so would raise the amount of "noise" for everyone around you. In turn, they would start speaking louder also and further escalate the amount of noise and interference. Before you know it, everyone in the room is only able to communicate from a few feet away from each other (Figure 5-2). If you have ever lost your voice at a rowdy party, or had to lean in to hear a conversation, then you have firsthand experience with SNR.

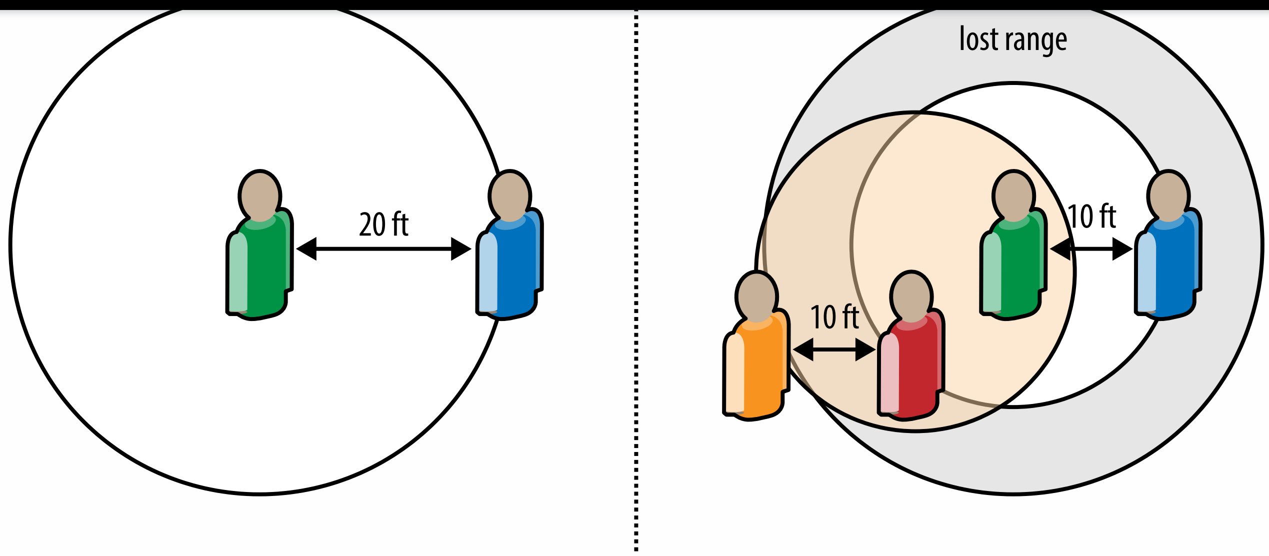


Figure 5-2. Cell-breathing and near-far effects in day-to-day situations

In fact, this scenario illustrates two important effects:

Near-far problem
: A condition in which a receiver captures a strong signal and thereby makes it impossible for the receiver to detect a weaker signal, effectively "crowding out" the weaker signal.

Cell-breathing
: A condition in which the coverage area, or the distance of the signal, expands and shrinks based on the cumulative noise and interference levels.

One, or more loud speakers beside you can block out weaker signals from farther away—the near-far problem. Similarly, the larger the number of other conversations around you, the higher the interference and the smaller the range from which you can discern a useful signal—cell-breathing. Not surprisingly, these same limitations are present in all forms of radio communication as well, regardless of protocol or underlying technology.

Modulation §

Available bandwidth and SNR are the two primary, physical factors that dictate the capacity of every wireless channel. However, the algorithm by which the signal is encoded can also have a significant effect.

In a nutshell, our digital alphabet (1's and 0's), needs to be translated into an analog signal (a radio wave). *Modulation* is the process of digital-to-analog conversion, and different "modulation alphabets" can be used to encode the digital signal with different efficiency. The combination of the alphabet and the symbol rate is what then determines the final throughput of the channel. As a hands-on example:

- Receiver and sender can process 1,000 pulses or symbols per second (1,000 baud).

alphabet (e.g., 2-bit alphabet: 00, 01, 10, 11).

- The bit rate of the channel is 1,000 baud × 2 bits per symbol, or 2,000 bits per second.

The choice of the modulation algorithm depends on the available technology, computing power of both the receiver and sender, as well as the SNR ratio. A higher-order modulation alphabet comes at a cost of reduced robustness to noise and interference—there is no free lunch!

Note

Don't worry, we are not planning to dive headfirst into the world of signal processing. Rather, it is simply important to understand that the choice of the modulation algorithm does affect the capacity of the wireless channel, but it is also subject to SNR, available processing power, and all other common trade-offs.

Measuring Real-World Wireless Performance §

Our brief crash course on signal theory can be summed up as follows: the performance of any wireless network, regardless of the name, acronym, or the revision number, is fundamentally limited by a small number of well-known parameters. Specifically, the amount of allocated bandwidth and the signal-to-noise ratio between receiver and sender. Further, all radio-powered communication is:

- Done over a shared communication medium (radio waves)
- Regulated to use specific bandwidth frequency ranges
- Regulated to use specific transmit power rates
- Subject to continuously changing background noise and interference
- Subject to technical constraints of the chosen wireless technology
- Subject to constraints of the device: form factor, power, etc.

All wireless technologies advertise a peak, or a maximum data rate. For example, the 802.11g standard is capable of 54 Mbit/s, and the 802.11n standard raises the bar up to 600 Mbit/s. Similarly, some mobile carriers are advertising 100+ MBit/s throughput with LTE. However, the most important part that is often overlooked when analyzing all these numbers is the emphasis on *in ideal conditions*.

What are ideal conditions? You guessed it: maximum amount of allotted bandwidth, exclusive use of the frequency spectrum, minimum or no background noise, highest-throughput modulation alphabet, and, increasingly, multiple radio streams (multiple-input and multiple-output, or MIMO) transmitting in parallel. Needless to say, what you see on the label and what you experience in the real world might be (read, will be) very different.

Just a few factors that may affect the performance of your wireless network:

- Amount of background noise in current location
- Amount of interference from users in the same network (intra-cell)
- Amount of interference from users in other, nearby networks (inter-cell)
- Amount of available transmit power, both at receiver and sender
- Amount of processing power and the chosen modulation scheme

In other words, if you want maximum throughput, then try to remove any noise and interference you can control, place your receiver and sender as close as possible, give them all the power they desire, and make sure both select the best modulation method. Or, if you are bent on performance, just run a physical wire between the two! The convenience of wireless communication does have its costs.

Measuring wireless performance is a tricky business. A small change, on the order of a few inches, in the location of the receiver can easily double throughput, and a few instants later the throughput could be halved again because another receiver has just woken up and is now competing for access to the radio channel. By its very nature, wireless performance is highly variable.

Finally, note that all of the previous discussions have been focused exclusively on throughput. Are we omitting latency on purpose? In fact, we have so far, because latency performance in wireless networks is directly tied to the specific technology in use, and that is the subject we turn to next.

« Back to the Table of Contents