Everything You Always Wanted to Know About 5G

(but were afraid to ask)

by Geoff Hollingworth & Peter Christy

Mobiledge X

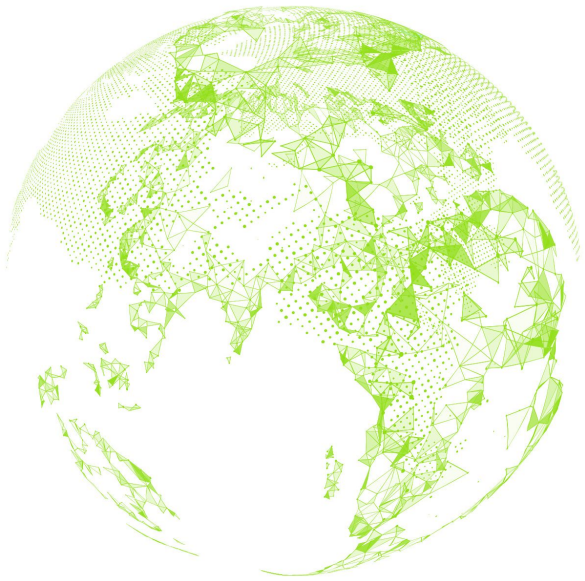
You would have to live in a cave not to know that “5G” is coming with wondrous new cellular capabilities. That said, for most people it’s still unclear exactly how 5G will change your life, especially if your background is with the Internet and Cloud, not the global mobile infrastructure. We hope this note makes the impact a little bit clearer. The most important challenge for 4G/LTE and 5G alike is simply building out a lot of new capacity. With 5G, individual connections are faster, and there are improvements to reduce communication latency and improve power efficiency.

That’s the short answer...

version 1.2 (August 2019)

The Global Mobile System is Pretty Amazing

The global cellular system is an outstanding example of what consumer technology can and should be – it works on a global scale, it provides both technical and business services – you can make a call from anywhere in the world, you get charged for it, everyone involved gets paid – and does so transparently and affordably. And you don't have to find a recent MIT graduate to help you set it up. It's a remarkable accomplishment.



It's all the more amazing considering that delivering this global service requires the coordinated technical and business interoperation of multiple completely independent service operators, and that mobile devices have been standardized so that there is a broad set of phone choices you can make from many independent vendors. If that weren't enough, the mobile operators can buy compatible equipment from multiple equipment vendors as well. In other words, compared to the Internet, the global mobile infrastructure is more complex technology, is a business system as well as a networking, and provides a greater degree of standardization.

All of this compatibility and interoperability is achieved through a set of comprehensive international standards that end up being quite different in nature from the standards that define the Internet[1], The comprehensive nature of cellular standards, and the process of implementing and evolving those standards creates the generational cadence that characterizes the evolution of the cellular infrastructure – it happens in big, lengthy and costly steps. We don't have comparable formalized generations on the Cloud or the Internet. These generations define the operation of the entire global mobile ecosystem in a way that you have to understand if you want to understand what "5G" is.

[1] Mobile standards are defined as a comprehensive whole (e.g. what constitutes 5G). The Internet standards are defined for each protocol. The standards evolve independently of one another; a network device chooses which protocols to support and at one revision level, and can choose to run advanced standards for one protocol, and long-standing for another. All 5G elements are mutually compatible. An Internet network engineer has to carefully assure that the various devices and revision levels are mutually compatible. The two approaches are very different; each as its advantages.

"5G" and Mobile

"5G" is broadly understood to be the next cellular "generation," but beyond that it's pretty hard for most outside the industry to unpack exactly what benefits accrue: what applications get better (and why), what new applications are enabled (and how)? Depending on what your role is in the cloud/Internet commercial ecology, why would you care about 5G?

A word of caution: the cellular infrastructure is nothing like the Internet, if that's what your background is. Telephony systems have always been purpose-built for, you guessed it, telephony. In contrast, when the Internet protocols were first created it was decided that any application-specific network optimization (e.g., telephony, video, transaction processing) would be done on top of a universal basic transport system (TCP/IP). The cellular infrastructure isn't a simple network like the Internet, it's better understood as a complex, distributed application system with integrated user and service management. 5G isn't just new network technology (new RFC's in Internet terms); practically it's much more.

The Internet architecture is pretty simple. The Internet consists of millions of routers, each connected to one-or-more neighbor routers. That's all there is: there is no master controller; there is no central operations center watching over everything; there is no standardized business model or practice. Each router notices when its local environment changes, for example, it sees a new device or link, or a device or link disappears. When this happens the router sends notice of the change to its neighbors. Over time these routing update messages propagate as far as they have to at which point the Internet will have adjusted to the change.

The global mobile system is totally different in many ways. First of all, it consists of a number of distinct subsystems, each quite complex. There is a Radio Area Network that decides at any moment what cell (specific radio channel) should be used by each device needing service, and adjusts all that dynamically as load changes and as devices move around.

The radio system converts the radio signals into a digital stream that includes a voice component and a data component. The mobile system started as a voice system and is now predominantly a data system (in terms of the payload and capacity driver). The voice system connects a mobile users to other telephones, whether they be mobile or wired. The data system provides connections between the device and Internet-located services, a quite different problem. 5G includes many changes to the radio system: many new frequency bands (including the use of non-licensed spectrum for the first time), new modulation and antenna systems that get more data in the allocated spectrum, and new radio protocols that reduce latency and increase battery life.

Next there is a substantial network system that provides the underlying transport under the voice and data systems and interconnects and coordinates the many different operators that collectively form the global mobile system. 5G includes important changes to the internal networking to make interconnectivity more efficient and to minimize the load imposed by data connections of the overall mobile infrastructure.

Finally, there is a complex, global business, authentication and payment system. Each mobile user has a business relationship with an operator that specifies what they can do the rate at which they will be charged for each service, and bills the user for services rendered and collects payments. The mobile operators, in turn, all have business relationships with other operators, that define how resources will be shared and how operators will be compensated. That underpins a quite amazing, global "roaming" system that provides cellular users service beyond the reach of the cells that operator has, and does so a local, regional and global basis. When you turn on a phone in another operator's territory, what you are authorized to do is quickly understood, you're provided those services, and the foreign operator is compensated for providing those services. 5G will make it easier to take advantage of the unique characteristics of the mobile infrastructure within applications run on mobile devices.

Most of the 5G discussion is at the radio level: faster data connections, lower latency, more frequencies to use, smaller cells, longer service for battery powered devices. There are also important changes at the networking level aimed at connecting to the Internet more quickly (further reducing latency on those connections, and enabling efficient access to computational services near the user (at the "edge").[2]

[2] When 5G was being conceived there were parallel discussions about what the Internet of the future would be like, including the notion that the Internet would evolve to be based on abstract content-based addressing rather than the physical addressing that IP is based on. Such changes would have been important since physical addressing makes mobility more difficult. No such large changes have come to pass however.

Why is 5G So Confusing to Outsiders

The confusion about 5G stems in part from its necessary multiple personalities, and in part from the enormous role of standards.

- 5G is a next generation technology standard, in an industry where standards play a critical role. Long before any new generation gear is sold the technology standards have to be specified in detail; the new functionality has to be imagined and defined long in advance of bringing it to market.

- At the same time, 5G (any new generation) is necessarily a marketing promise. In an industry where progress is so expensive (Trillions of dollars on a global scale) "build it and they will come" just doesn't work. Mobile operators need to see the demand before they will invest in a new generation of technology. For equipment suppliers, there is a lot at stake in terms of future revenues.

With the Internet and the Cloud, innovation is ongoing and incremental: add something and see who uses it for what. The cellular industry is fundamentally different: progress occurs in big, long, expensive steps – generations – and much more of it is standardized (including parts of the business systems). The cellular industry has to decide what comes next many years before it hits the market; that's completely different from the Internet and Cloud.

Refreshing the infrastructure in order to move to the next generation is expensive – it's estimated that nearly $2bn was invested in the global cellular infrastructure in the last decade. Implementing 5G will be costly, and financially scrutinized mobile operators won't invest until they believe the return is realistic. So the mobile industry has to start beating the drum early on and evangelize what's coming and create real demand for it, in many cases well before the world at large understands why they need it, or it will never be implemented.

5G as a technology standard is very concrete, specific and formally defined. But much of the public discussion is about the marketing promise and is aspirational more than factual. To make sense out of it you have to keep the two separate in your mind. We know what 5G is, but we don't yet understand exactly how we're going to use it, or how and where it will create value. That's the nature of cellular generations, and that's why it's hard as an outsider to sort out 5g – what's the very concrete technology and standards are one thing; exactly how that will play in the market something entirely different.[3]

[3] The history of 4G is a great example. The intent of 4G was to change the internal networking of the mobile infrastructure from circuit switched to packet switched. Packet switching enabled incremental improvements like LTE

Context

A few more details will help make the 5G changes understandable.

The Miraculous Decade

For the last 60 years we've been on a remarkable ride driven by continuing progress in semiconductor technology ("Moore's Law"). And then a set of "independent" events occurred a little more than a decade ago, and the world of IT changed rapidly as a result. The pivotal event was the introduction of the iPhone: a compelling vision of what a mobile, personal computer could be – after that we all knew we wanted one for personal and business use; the world of "personal computers" was changed forever. The second event was Amazon's introduction of Amazon Web Services that provided virtual system resources, on demand, where you only pay for what you use – after that new, software-centered, companies didn't need to buy their own servers, nor necessarily take VC investment in order to do that; the world of IT entrepreneurship was changed forever. The third parallel event was the 4G cellular generation.

The Role of 4G/LTE

And then there was 4G/LTE, the previous cellular generation that started to be built out at more of less the same time and has provided the broadband cellular data services we all use today and that to a large degree define the value of the smartphone – e.g., high-definition video communications with friends, or watching videos while commuting to work. With 4G/LTE we all know why we want high-bandwidth cellular connectivity and have a good idea how we will use it.

Getting back to cellular generations and the role of demand, here's the ironic part: 4G wasn't motivated by broadband data connectivity, it just enabled it. The "purpose" of 4G was to change the networking architecture under the global mobile infrastructure from circuit to packet switching. The very fortunate side effect was enabling the development and buildout of broadband data, which in the end has been the hallmark of 4G/LTE.

This ironic history is important to understand because it explains the issues in 4G data connectivity that are limitations going forward that 5G is addressing:

- Latency: the latency in cellular data is quite noticeable (due to the use of forward error correction if you have to know). 5G makes important changes in radio protocols.

to provide broadband data connectivity, the value that drove and paid for the rapid buidout of 4G/LTE. Broadband data connectivity wasn't at the center of the a priori promise at all.

- Interconnecting to the Internet: This is the part that's a real mess – for reasons that make historical sense, a cellular data connection bridges to the Internet at a network gateway that doesn't make a lot of sense from an Internet perspective, and, importantly, maybe distant from the tower the subscriber is using, creating challenges to adding "edge" services.

An astute reader may have noticed that "insufficient bandwidth" isn't among the problems listed. That's because a good 4G/LTE data connection (100Mbps) is pretty sufficient for most mobile needs. More bandwidth is better for many reasons, of course.

Edge Computing

While all this has been occuring, a separate Internet discussion has started on "edge" computing. The basic ideas are simple: what do you do if you need to put computation or services nearer to the user (or device) than is possible with today's public cloud (where "closer" is defined as a "better" network connection – lower latency or greater bandwidth, for example), especially if you want to preserve the benefits of the cloud (on-demand, pay only for what you use)?

As more and more Internet use moves to wireless, mobile devices, the question of edge computing in the mobile infrastructure comes up. As we'll see later, the existing mobile infrastructure and design poses challenges doing that.

The Challenges of Radio Connectivity

Radio data links are complex (lots of different technical and regulatory issues) but at the same time simple, more capacity requires more and smaller cells. Most readers won't want to understand the technical issues in detail (although they will delight your inner engineer) but everyone can and should understand the important bit – more cells,smaller cells.

One might imagine that sending data via radio signals through the "ether" would be much easier and more effective than through wires or optical fibers, given their physical imitations, but it's really not the case. The amount of data you can squeeze into a radio channel is theoretically limited by the radio spectrum you are able to use you, the modulation technology you use,and the noise encountered (competing radio signals from many sources). Furthermore, as a radio signal propagates it diminishes in strength more rapidly than on wires or optical fibers, so the data capacity depends strongly on the distance between sender and receiver.

If that weren't confusing enough, all frequency bands are not equal. Today's cellular systems are built from the bits and pieces of spectrum that can be licensed (dedicated in some regions to a single purpose). With 4G/LTE, these radio bands range from about 700MHz to 3,000MHz (3GHz). Even within this range, there is a big difference in the ability of the radio signals to penetrate a building or even a sunscreen-coated glass window, depending on the frequency. From a subscriber's perspective, the lower frequencies are best because the penetration (usability within buildings) is better. Frequency and penetration is an important issue in 5G because many of the newer frequencies that are being made available are at much higher frequencies (up to ~30GHz) with even worse penetration.

To summarize, there are no radio "silver bullets" (or even brass bullets for that matter) that would enable 5G changes by themselves to solve the mobile broadband demand problem. A little more data can be squeezed through existing radio spectrum (and 5G includes such improvements), and new spectrum can be added (which 5G does as well), but, in the end, to get more capacity you have to add a lot more cells, with each cell being smaller (subject to less interference). If you want to be a cellular data expert just remember that the *sine qua non* (the essence) of the capacity solution going forward is a lot more small cells (4G, 5G, WiFi). The challenge is not exotic radio advances – it's connecting all those new cells back into the Internet,. The obvious way to connect them back in is with fiber optics, and that can require trenching city streets and installing them in existing buildings, neither of which is interesting tech, just hard, costly and time-consuming work.

5G and WiFi

5G and WiFi are competitive approaches to providing additional wireless capacity. 5G advocates will say (correctly) that cellular radio systems are superior to WiFi. Like all the rest of the Internet, WiFi is based on IP packet networking and Ethernet physical networking, and for that reason WiFi isn't "managed" at the physical/radio level whereas cellular radio is. If you want to use a cellular link, you make that request to a controller and are allocated capacity. If you want to use a WiFi link you wait until you think it is free and try to transmit. In the most demanding circumstances (e.g., a factory floor with clustered, radio-connected devices), 5G has superior performance (e.g., greater capacity utilization, lower latency).

That does not, however, mean that 5G will greatly erode the WiFi market. For less demanding applications. the technologies perform comparably. More importantly, for an enterprise that has adopted Ethernet/IP networking architecture broadly (e.g., a typical Cisco enterprise customer) there will be strong and meaningful resistance to adding a completely different technology that requires new skills and training. After all, Cisco has spent the last 30 years replacing different network technologies with Ethernet/IP alternatives, and by doing so, reduced the cost and increased the availability of enterprise networking. It was never because Ethernet/IP was better than what was replaced – it was just good enough.

Data Architecture

The internal data networking in mobile infrastructure needs some serious fixing. At a high level, the problems and solutions are easy to describe. That's my goal here. The solutions are specific to the details of the mobile infrastructure. Here are the key points:

- Broadband data is very much an afterthought in the design and implementation of the global mobile infrastructure. Broadband connectivity was a wonderful side-benefit of how 4G turned out, but not the intent. The data connection design was just fine for a voice network with a little bit of data; less so for a service increasingly dominated and driven by data requirements.
- Mobility was equally not part of the design of TCP/IP and hence the Internet. The Internet functions adaptively and brilliantly to connect two physical endpoints in light of the internal, ongoing, changes to the Internet, but it is fundamentally designed to get data to the connected router that is closest to you. If you move to a different physical location the impact on Internet routing is broad and ugly, and it takes a lot of time to adapt.

When data connectivity was added to the global mobile telephony infrastructure, this conflict between the physical Internet and cellular mobility was addressed simply and elegantly by assigning each connected mobile user to a data gateway – where the mobile infrastructure connected to the Internet – that was unlikely to change. For many years the AT&T network in the US had two such gateways – one in the east and one in the west – unlikely to change as your drove around.

This architecture was very compatible with TCP/IP and the Internet. It added some latency to cellular connections compared to a gateway closer to the mobile user that let the Internet find a closer path to a specific website (for example, a local service for example), but remember that 4G connectivity has considerably latency to begin with.

Moving forward the goal is clear–separate a data connection from the cellular connection as soon as possible. Doing that reduces the load on the cellular infrastructure, it enables a shorter connection to a local Internet resources and it enables the addition of edge services that are as near to the cell tower in use as possible.

The fix sounds simple but is how to do this without causing problems with TCP/IP routing and without destabilizing the networks in place that support the cellular infrastructure (It's a much longer discussion, but making changes to wide-area-networks is difficult because changes have to be made to the configuration of many, distributed devices, including ones in unmanned ("lights out") locations.)

Summary

5G is coming soon to a mobile operator near you, if it hasn't already arrived. 5G connections can be faster and better (lower latency) but that won't make much difference to applications that already work but will enable more demanding applications and new categories such as IoT and VR/AR.

5G architecture changes will help mobile operators deal with ever increasing volumes of cellular data which will keep infrastructure costs down while capacity is added, and enable more effective cloud computing, including cloud services at the edge of the mobile network.

The most important challenge facing wireless network operators is simply adding huge amounts of capacity to keep up with growing demand. Doing that requires adding a lot of new, smaller cells and that requires wiring all those new cells (4G, 5G or WiFi) back into the rest of the Internet.

5G is the important next version of the global cellular infrastructure. There is no doubt that mobile infrastructure will play in increasingly important role in our transformation use of the Internet and the Cloud; we won't know what the all exciting new applications and services are until they're invented, nowever.