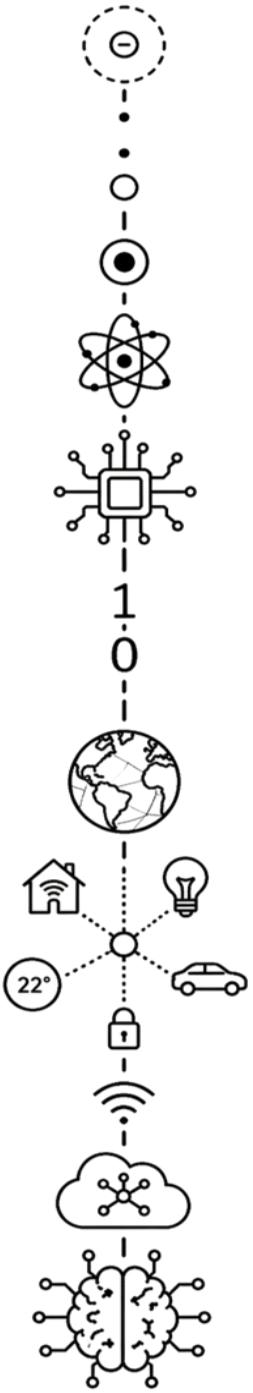


PREVIEW
full edition available
for purchase at
www.ardiri.com

Inside The Mobile Network

Understanding how Mobile Communications Work — From 1G to 5G and the standards that connected the world



Aaron Ardiri

© 2026 Aaron Ardiri
All rights reserved.

First Edition - 2026

ISBN 978-91-535-1698-9

Dedicated to **Tomas Bennich.**

For as long as I have known you, you have had a remarkable ability to bring people together—creating introductions, starting conversations and building connections across Sweden's mobile and technology community.

Over the years, many of those connections have become friendships, collaborations and opportunities that would never otherwise have existed. A mobile network is ultimately about connecting people. It therefore seems only fitting that this book is dedicated to someone who has spent so much of his life doing exactly that.

FOREWORD

I spent thirty-five years at Ericsson, during a period in which mobile communication changed almost beyond recognition. What began as a technology primarily concerned with connecting people through voice evolved into a global infrastructure carrying data, applications, media and, increasingly, communication between machines. I had the privilege of experiencing parts of that transformation from close range, working both technically and commercially with technologies and standards that would eventually become part of the mobile networks we now take for granted.

Some of my most memorable years were spent in Japan, working closely with Softbank and NTT DoCoMo during an extraordinary period in the development of mobile communications. Japan was often ahead of much of the world in exploring what mobile networks could become, and the relationship between operators, equipment vendors, handset manufacturers and standards was both technically challenging and commercially fascinating. Later, after returning to Sweden, my attention increasingly turned toward the Internet of Things and the question of how the principles developed for connecting billions of people might be extended to a world containing potentially far more connected machines.

It was through that interest that I first met Aaron in 2018.

Our initial discussions concerned RIoT Secure and its approach to IoT technology, but they quickly became much broader. One expression from those early conversations has stayed with us ever since: the Russian matryoshka doll. Open one layer of a technical system and another appears inside it. Open that one and there is another. What initially looks simple from the outside is usually the product of many interacting technologies, protocols, architectural decisions and engineering compromises hidden beneath the surface.

That observation applies particularly well to mobile networks.

To most people, a mobile network is represented by a few bars at the top of a screen. A phone is switched on, a network name appears and communication begins. Very little of the extraordinary machinery required to make that possible is visible to the person using it. Radio systems discover one another. Identities are established. Subscribers are authenticated. Mobility is managed. Sessions are created. Packets are routed. Calls cross networks and national borders. Thousands of independently developed products interoperate because engineers and companies agreed, sometimes after considerable debate, on precisely how they should behave.

During my years at Ericsson, I spent considerable time working with 3GPP standards, both from technical and commercial perspectives. Standards documents are necessarily precise. They define interfaces, procedures, messages and behavior so that systems created by different organizations can work together. What they do not necessarily provide is the story of why the system became what it is. That distinction is one of the things I appreciate about Inside the Mobile Network.

For someone who participated in parts of this evolution, reading the material inevitably brings back memories. Names of network elements, interfaces and standards that may appear historical today once represented active engineering problems, commercial decisions and sometimes vigorous discussions between people and organizations with very different priorities. Technologies that now fit neatly into diagrams did not arrive neatly. They evolved through proposals, experiments, compromises, deployments and lessons learned from systems already operating in the field.

Aaron and I have approached technology from different directions, which is perhaps why we have enjoyed working together. Over the years since that first meeting, we have continued to exchange ideas across mobile networks, IoT, embedded systems, security, software architecture and many other subjects. Those conversations have frequently involved challenging one another's assumptions, asking why something was designed in a particular way and following another Russian doll when the answer revealed a deeper layer.

That same curiosity runs throughout this book.

The final chapters step back from the individual technologies to reveal the mobile network as a complete system, raising a question familiar to anyone who has worked within the industry: **how did we ever agree on all of this?** The answer leads into standardization and ultimately toward 6G, where many of tomorrow's decisions are still being made today.

After thirty-five years at Ericsson, I have learned that successful telecommunications engineering has an interesting characteristic: when it works well, most people never need to know that it exists. Enormous technical complexity disappears behind an experience as simple as taking a phone from your pocket. This book opens that experience and looks inside—like the Russian matryoshka doll Aaron and I discussed when we first met, there is always another layer waiting underneath.

John Fornehed

https://www.linkedin.com/in/johnfornehed/

October, 2026

Table of Contents

About the Author..... 1

The Analog Foundations..... 3
Before Cellular.....4
1G: The Analog Generation..... 28

The Digital Revolution..... 56
2G: Going Digital.....57
More Than Voice..... 90

The Mobile Internet..... 131
2.5G: The Internet Arrives.....132
3G: Mobile Broadband..... 167
4G: Everything Becomes IP..... 211

The Modern Mobile Network..... 272
5G: A New Architecture.....273
The Mobile Market.....344
Inside the Mobile Network.....430

Standardization and the Future..... 475
Standardization and the Road to 6G.....476

Appendixes..... 522
Mobile Network Generations at a Glance.....523
3GPP Release Timeline..... 524
3GPP Specification Families.....525
Acronyms and Network Elements..... 526

About the Author



Aaron Ardiri

https://www.ardiri.com/

I am a software engineer, systems architect, entrepreneur and technology leader with more than three decades of experience designing software systems spanning mobile computing, embedded devices, networking, cybersecurity, cloud platforms and the Internet of Things.

My involvement with mobile technology began during the formative years of handheld computing and grew alongside an industry undergoing extraordinary change. I have developed software across generations of mobile platforms, worked with device manufacturers and telecommunications companies, and spent much of my career exploring the systems that allow increasingly sophisticated devices to communicate reliably and securely.

Throughout my career, I have held technical leadership positions ranging from startup founder to Chief Technology Officer and Principal Software Architect, working with organizations including BlackBerry, Ericsson and numerous technology ventures across Europe and Australia. I am the founder and CEO of RIoT Secure AB, where I have led the development of secure device lifecycle management platforms and communication technologies for connected embedded systems.

Working in the mobile industry provided a perspective extending well beyond the handset itself. Devices are only the visible edge of an enormous global system involving radio technology, spectrum, base stations, core networks, subscriber identity, security, roaming, protocols and standards. Technologies developed by competing manufacturers and operators must work together across national borders and across generations of equipment.

Inside the Mobile Network grew from a desire to explain that system as a connected whole.

The book follows the evolution of mobile communications from early cellular networks through 1G, GSM and the emergence of mobile data to 3G, LTE and 5G. It explores not only how the technology works, but how the architecture evolved and how organizations such as 3GPP enabled thousands of engineers and companies to build interoperable networks used around the world.

At its heart, however, the book is driven by the same question that has followed me throughout my engineering career:

How does it really work?

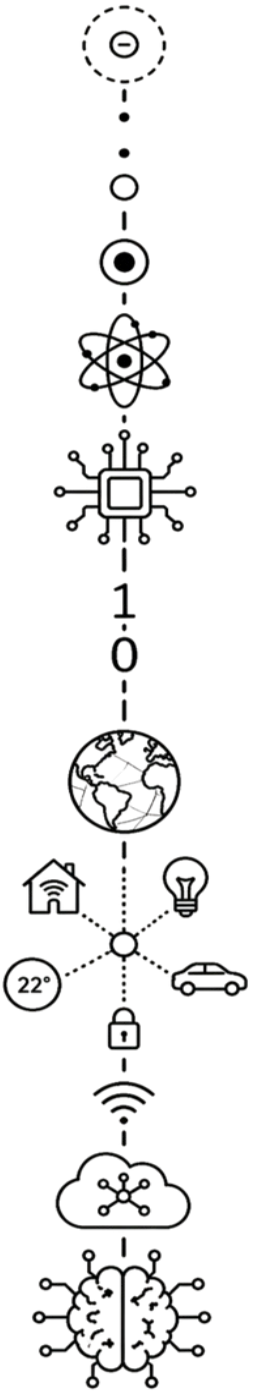
That question also lies behind my previous book **Inside the Machine**, which explores the representations, execution mechanisms and architectural ideas underlying modern computers, and my autobiography, **Technology, Entrepreneurship and the Reality of Building Systems**, which tells the personal story behind a career spent building technology.

Inside the Mobile Network extends that curiosity beyond the computer itself. Instead of looking inside a single machine, it looks inside one of the largest distributed systems humanity has built—a network that allows billions of devices to find one another, authenticate, communicate and remain connected while moving around the world.

I continue to design systems, explore new technologies and share my passion for engineering through writing, speaking and mentoring. I believe the most valuable skill an engineer can develop is not memorizing a particular technology or standard, but cultivating the curiosity to understand why systems are designed the way they are.

If this book encourages you to look beyond the signal bars on a mobile phone and see the extraordinary collection of engineering decisions, protocols, standards and infrastructure working behind them, then it has achieved exactly what I hoped.

ACT 1:
The Analog Foundations



Before Cellular

For most of human history, communicating with someone far away required either the message or the person to travel. You could meet face to face. You could write a letter and entrust it to someone capable of carrying it to its destination. The journey might take days, weeks, or even months, but the underlying relationship was simple: distance could only be overcome by physically moving something through it.

Telegraphy changed that relationship dramatically. A message could travel across great distances without a person carrying it from one place to another. Telephony went further still, allowing two people separated by distance to speak and hear one another almost immediately. But there was a condition hidden inside this new freedom.

The telephone had a place.

A home telephone belonged to a house. An office telephone belonged to a desk. A public telephone belonged to a street corner, railway station, hotel, or airport. The network could connect people across enormous distances, but to participate in that network they normally had to go somewhere a physical connection already existed.

Communication had conquered distance without conquering location. That distinction is easy to overlook today. We carry devices that allow us to communicate while walking through a city, sitting on a train, driving along a highway, or traveling through another country. The network follows us so naturally that being tied to a particular place now feels like the exception. For early telecommunications engineers, however, separating communication from location presented an entirely different problem.

Removing the wire meant using radio, but radio introduced constraints that the fixed telephone network did not have to solve in the same way. Spectrum was finite. Transmissions could interfere with one another. Radio coverage changed with distance and geography, and increasing transmitter power could extend coverage without necessarily allowing more people to communicate at the same time.

Then there was movement itself. A fixed telephone network could associate a connection with a known physical line. A mobile user might be in one place when communication began and somewhere entirely different a few minutes later. If communication was truly to become mobile, the network would somehow have to accommodate a participant whose location was continuously changing.

The cellular network did not appear as the immediate answer.

Long before engineers divided landscapes into cells, allocated frequencies for reuse, or developed mechanisms for handing a connection from one radio site to another, people were already experimenting with ways to extend telephone service beyond the wire. Those early systems made mobile communication possible—they also revealed why something better was needed.

Early Mobile Radio and Radio Telephony

Long before anyone carried a telephone in a pocket, there was already a desire to communicate with people who were moving.

For most of the history of the telephone, however, movement and telephony were almost contradictory ideas. A telephone belonged to a place. It sat on a desk, hung on a wall, or rested on a table because it was physically connected to the telephone network by a pair of wires. The network knew how to reach the telephone because the telephone stayed where the wires ended.

Radio changed that relationship.

A radio transmitter did not need a continuous physical connection between two locations. Information could be carried through electromagnetic waves, allowing a receiver to hear a transmission from kilometers away. Ships could communicate with shore stations. Military units could coordinate across a battlefield. Police officers, emergency services, taxi drivers, and other mobile workers could remain in contact while traveling through an area.

Yet these early radio systems were not telephone networks without wires. Most were designed around a very different model of communication. A radio channel might be shared by many users, with only one person speaking at a time. Pressing a button activated the transmitter; releasing it allowed someone else to respond. Everyone using the same channel might hear the conversation.

That was perfectly adequate for many applications. A dispatcher did not necessarily need a private connection to every taxi. A police radio system benefited from several officers hearing the same message. What mattered was the ability to communicate with people who were away from a fixed telephone.

A telephone call created a different expectation.

When someone lifted a telephone handset and dialed a number, they expected a private conversation with one particular person. The telephone network established a connection between the two endpoints and maintained it for the

duration of the call. Bringing that experience into a moving vehicle meant connecting two very different systems: a radio system designed to communicate through the air and a telephone network designed around fixed subscribers and switched connections.

Early radio telephone systems attempted to build that bridge.

A vehicle could contain a radio transmitter and receiver that communicated with a fixed radio station covering a large geographic area. That station, in turn, provided access to the public telephone network. A person in the vehicle could therefore communicate with someone using an ordinary telephone even though there was no wire connecting the vehicle itself.



Figure 1.1. *A mobile radio communicates with a fixed radio station, which provides the connection into the public telephone network.*

The arrangement contains the essential ingredients of mobile telephony in a surprisingly recognizable form. There is a mobile device, a radio connection, fixed infrastructure, and a larger network capable of reaching other subscribers. Modern mobile networks still contain equivalents of each of these elements, although almost everything about their implementation has changed.

What early systems lacked was scale.

A radio transmitter covering a city might reach vehicles across a very large area, which sounds desirable. Fewer radio sites were needed, and a sufficiently powerful transmitter could provide impressive geographic coverage. But every conversation required access to a radio channel, and the number of channels available was limited. Once those channels were occupied, another subscriber could not simply begin another conversation. This reveals one of the most important ideas in the history of mobile communications:

Coverage and capacity are not the same thing.

A system may be capable of reaching thousands of vehicles while still supporting only a small number of simultaneous conversations. Making the transmitter more powerful could extend the distance over which the signal could be received, but it did not create additional radio spectrum. In some circumstances, greater range could make the problem worse by preventing the same frequencies from being reused elsewhere without causing interference.

> 🔥 **Engineering Insight — Wireless Communication**
>
> Wireless communication introduces a constraint that fixed telephone users rarely need to consider: everyone is sharing the same physical medium. A wire can be installed for another subscriber. Radio spectrum cannot simply be manufactured when demand increases.

That limitation would eventually become the central engineering problem of mobile telephony. If radio channels were scarce, then building a mass-market mobile telephone system could not depend on assigning a dedicated frequency to every subscriber. Nor could it rely indefinitely on a handful of powerful transmitters covering enormous areas.

The solution would require engineers to think differently about the relationship between geography, radio coverage, and the telephone network. Before that could happen, however, the radio equipment itself had to become mobile enough to take the telephone onto the road.

Vehicle-Mounted Telephone Systems

> **If a telephone could communicate by radio, an obvious question followed: why not put one in a car?**

The idea was compelling, but the technology of the time made the word *mobile* somewhat generous. Early radio equipment required substantial amounts of electrical power, transmitters were bulky, antennas needed to be large enough to operate efficiently, and the supporting electronics could occupy considerable space. A handheld telephone was not yet practical, but a vehicle provided something a person walking down the street could not: a battery, an antenna mounting point, and enough room to carry the equipment.

The automobile therefore became one of the first practical homes for the mobile telephone.

From the subscriber's perspective, the goal was deceptively simple. Someone sitting in a car should be able to communicate with a person using the ordinary telephone network. Achieving that required much more than installing a radio beside the driver's seat. The vehicle needed to establish a radio

connection with a fixed station, and somewhere beyond that station the radio conversation had to be connected into the existing telephone system.

Early systems often depended heavily on human operators to make that connection possible. A subscriber could contact a mobile telephone operator over the radio, provide the number they wanted to call, and wait while the operator established the connection through the public telephone network. An incoming call involved the reverse process: the system had to determine that someone was trying to reach a particular mobile subscriber and somehow alert that subscriber over the radio network.

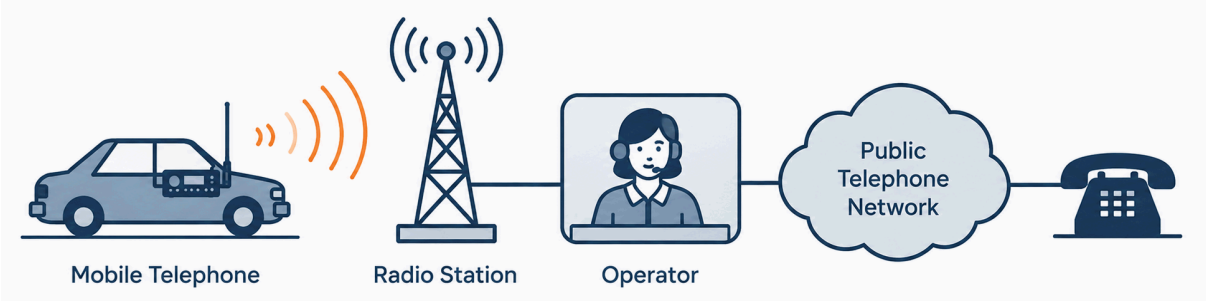


Figure 1.2. *An early vehicle-mounted telephone reaches the public telephone network through a radio base station and, in many systems, a human operator who establishes the call.*

This arrangement may seem primitive compared with a modern mobile network, but it exposed several problems that remain recognizable today. The network needed a way to identify a subscriber. It needed to know whether that subscriber was reachable. It needed radio resources over which communication could take place. It needed infrastructure connecting the radio system to other networks. And it needed some mechanism for establishing and terminating a call.

The difference was that much of the intelligence that would eventually become part of the network was still provided manually.

As systems evolved, automation began removing the operator from the process. Subscribers could dial calls themselves, and networks could perform more of the work required to establish a connection. The experience moved closer to that of an ordinary telephone, but the underlying radio limitations remained. A city might have many people interested in mobile telephone service while having only a small number of radio channels available for conversations.

The physical equipment imposed limitations of its own. Powerful transmitters consumed significant energy, which reinforced the dependence on vehicles. Antennas mounted outside the car provided better radio performance than

something carried inside it, while the vehicle's electrical system could supply power without requiring enormous portable batteries.

ⓘ **Did You Know? — Early Mobile Phones**
The earliest mobile telephones were not merely large versions of later handheld phones. The vehicle was effectively part of the system, providing the power, space, and antenna installation needed to make practical mobile radio communication possible.

This also helps explain why the earliest conception of a mobile telephone was quite different from the one we have today. Mobility did not initially mean that a telephone could accompany a person everywhere. It meant that the endpoint of the telephone network no longer had to remain at a fixed address. A subscriber could travel across a coverage area and remain capable of communicating with the fixed telephone network.

That was already a profound change.

But placing the telephone in a car solved only the physical problem of carrying the equipment. It did nothing to create more radio channels. If ten channels were available, only a limited number of conversations could use them simultaneously, regardless of whether ten people or ten thousand people wanted the service. The equipment could become smaller. Calls could become automated. Coverage could expand. The radio spectrum remained finite.

Spectrum Scarcity and Channel Capacity

Making mobile telephones practical enough to install in vehicles solved an important problem, but it exposed a much more difficult one. There were only so many conversations the system could carry at once.

Consider a city served by an early mobile telephone system with a small collection of radio channels. Each active telephone conversation required access to one of those channels. While the conversation continued, that radio resource could not simply be given to another subscriber in the same coverage area. Once every available channel was occupied, the system had reached its capacity. Another subscriber might have a perfectly good radio signal and still be unable to make a call.

This distinction is easy to overlook. We naturally associate a strong signal with a working mobile service, but radio coverage answers only one question:

> **Can the network and the mobile device hear one another?**

Capacity answers a different question:

> **Does the network have enough resources to serve everyone who wants to communicate?**

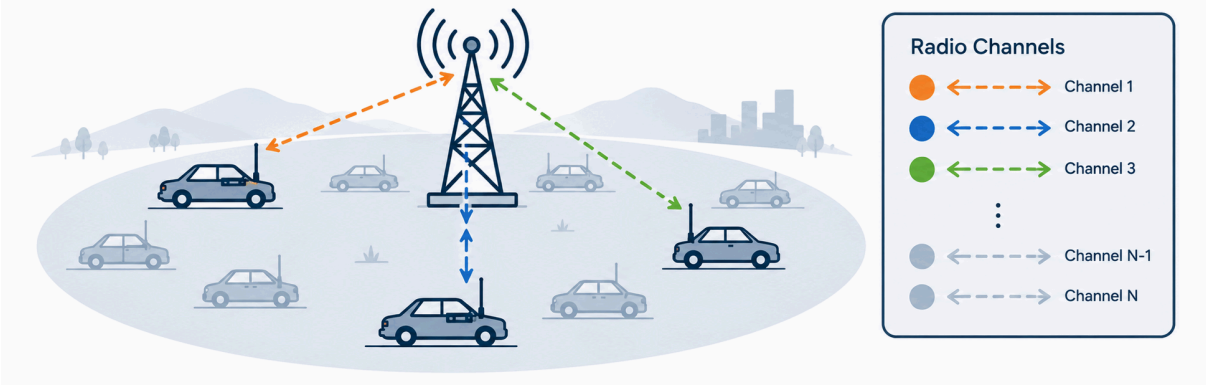


Figure 1.3. *A radio system may cover many subscribers, but the number of simultaneous conversations is limited by the radio channels available within that coverage area.*

Imagine, for simplicity, that a mobile telephone system has ten channels available across an entire city. Ten subscribers could potentially hold ten simultaneous conversations. An eleventh subscriber attempting to place a call would have to wait until one of those conversations ended and a channel became available.

Adding more subscribers would not add more capacity.

One hundred subscribers might be manageable if they rarely made calls at the same time. Ten thousand subscribers would create a much greater probability that all ten channels were already occupied. The network could sell service to more people than it could serve simultaneously because most subscribers were idle most of the time, but there was still a very real limit to how far that approach could scale.

The obvious solution might appear to be allocating more frequencies, but radio spectrum is not an unlimited resource. Different frequencies are used by broadcasting, aviation, navigation, military systems, emergency services, satellite communication, and countless other applications. Mobile telephone systems could operate only within spectrum allocated for that purpose, and even within those allocations the usable spectrum had to be divided into channels, each occupying part of that finite resource.

The problem was made harder by interference. Two nearby transmitters could not necessarily use the same frequency independently. If their signals reached the same receiver with sufficient strength, the receiver could have difficulty separating the desired transmission from the unwanted one. Frequencies

therefore had to be planned with consideration not only for how many channels existed, but also for where those channels could safely be used.

> 🔥 **Engineering Insight — Spectrum**
>
> Spectrum is unusual because using a resource in one place can affect whether the same resource can be used somewhere else. Mobile-network capacity therefore depends not only on how much spectrum is available, but on how intelligently that spectrum is reused.

This turns spectrum scarcity into an architectural problem.

A conventional telephone network could expand by installing more cables, adding switching equipment, and increasing the capacity of its infrastructure. Radio networks did not have the equivalent luxury of simply installing another piece of spectrum. Engineers could improve equipment and use the available frequencies more efficiently, but the underlying physical resource remained shared and finite.

That reality would shape virtually every generation of mobile technology that followed. GSM would divide radio resources across time and frequency, while later systems would introduce increasingly sophisticated ways of allowing many users to share spectrum. LTE and 5G would dynamically allocate radio resources across time and frequency with a flexibility that early mobile telephone engineers could scarcely have imagined, yet the underlying problem never disappeared.

A modern network may use vastly more sophisticated techniques, but it is still deciding how to share a finite radio resource among many users who may all want to communicate at the same time.

The early systems made the limitation particularly visible. A single radio site could cover an enormous area, but every subscriber within that area was effectively competing for the same limited collection of channels.

Extending the reach of the transmitter would allow the network to cover more people. It would not necessarily allow more of them to talk. That distinction between **coverage** and **capacity** would become one of the most important insights behind the cellular network.

Coverage Versus Capacity

A powerful radio transmitter seems, at first, like an obvious advantage.

If a transmitter can reach farther, the network can cover a larger area. Fewer radio sites are required, subscribers can travel greater distances while remaining within range, and infrastructure becomes simpler. For an early

mobile telephone system, increasing coverage could appear to be exactly what engineers wanted. But the spectrum problem reveals something counterintuitive. A larger coverage area can actually make it harder to serve more subscribers.

Imagine a single radio station positioned high above a city. With sufficient transmitter power and a well-placed antenna, its signal might reach vehicles many kilometers away. From the perspective of coverage, this is excellent. A large geographic area can communicate with one station.

Every subscriber within that area, however, is sharing the same collection of radio channels. If the station has ten channels available, extending its coverage does not turn those ten channels into twenty. It merely allows more potential subscribers to compete for the same ten resources. The network can reach more people without gaining the ability to support more simultaneous conversations.

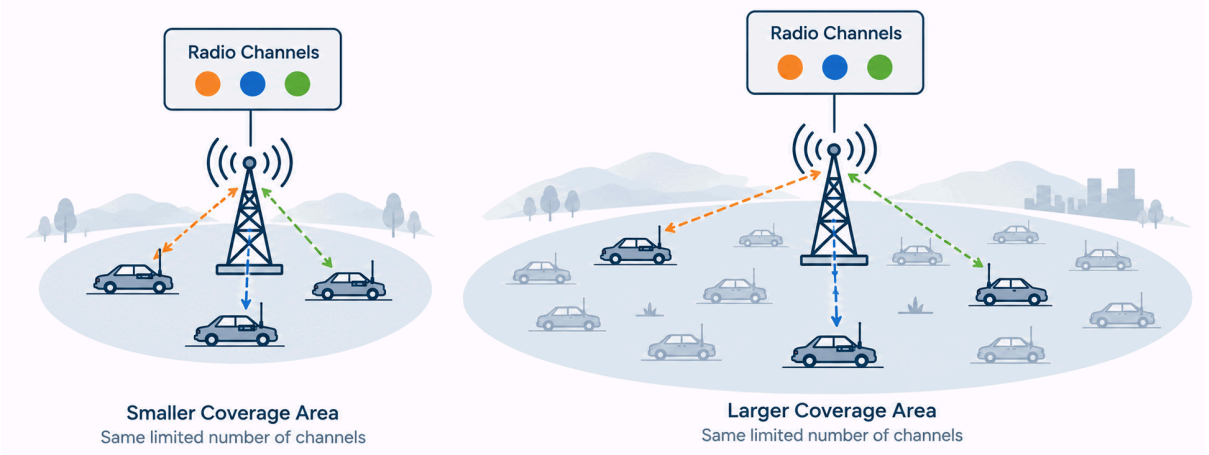


Figure 1.4. *Increasing the coverage of a single radio site allows it to reach more subscribers, but the same limited collection of radio channels must still be shared across the entire coverage area.*

This creates two different engineering objectives.

Coverage describes where the service is available. A network with greater coverage can reach subscribers across a larger geographic area. **Capacity** describes how much communication the network can support. A network with greater capacity can serve more simultaneous users or carry more traffic within an area.

The two are related, but improving one does not automatically improve the other. This distinction becomes especially important when considering transmitter power. Increasing power can make a radio signal usable at greater distances. That may solve a coverage problem, but it does not create additional spectrum. Worse, a powerful transmission continues traveling

beyond the subscriber it is intended to reach. If another radio system nearby attempts to use the same frequency, the two transmissions may interfere.

The very thing that improves coverage can therefore restrict reuse of the radio resource.

Consider two towns separated by some distance. If each uses low enough transmitter power that its signal remains largely within the surrounding area, both may be able to use the same radio frequency without causing unacceptable interference. Increase the power substantially, however, and the signal from one town may begin reaching receivers in the other. The frequency that could previously be used independently in both places has effectively become shared.

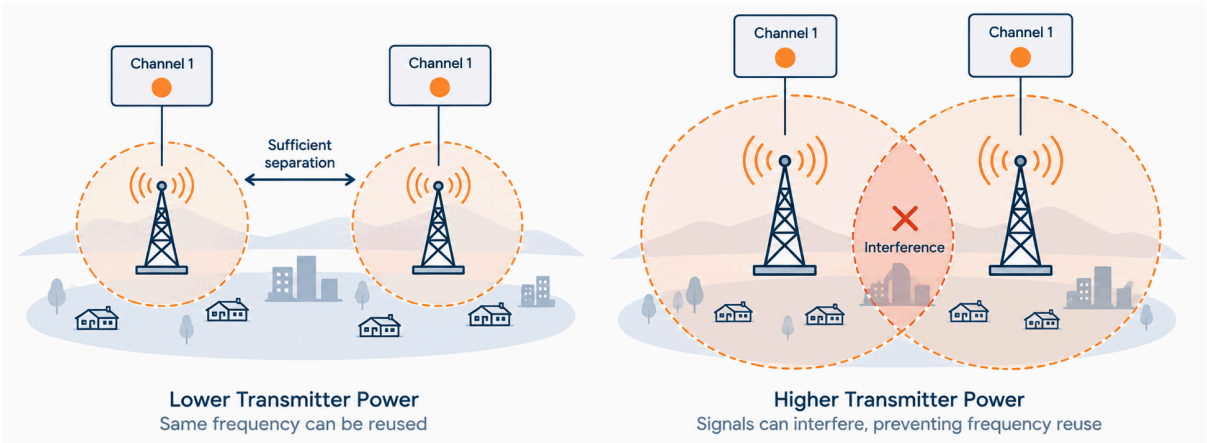


Figure 1.5. *Lower-power coverage areas can allow the same radio frequency to be reused at sufficient geographic separation, while excessive range can cause the signals to interfere.*

This changes the way we think about radio coverage.

The objective is not necessarily to make every transmitter reach as far as physically possible. A mobile network needs sufficient coverage to reach its subscribers, but it also benefits from controlling where signals are strong enough to matter. Limiting the geographic reach of a transmission can create opportunities to use the same frequencies again elsewhere.

ⓘ **Did You Know? — Transmitters**
Radio engineers often deliberately design systems so that individual transmitters cover less territory than they theoretically could. Restricting coverage may appear wasteful, but it can allow scarce spectrum to be reused elsewhere and dramatically increase the total capacity of the network.

The implication is profound. If one enormous radio coverage area forces everyone to share the same channels, perhaps the network should not be built

around one enormous coverage area at all. Instead of asking how powerful a transmitter must be to cover an entire city, engineers could ask a very different question:

What if the city were divided into many smaller coverage areas?

Each area would require its own radio equipment, making the infrastructure more complicated. The network would somehow need to coordinate those sites. A moving subscriber might leave one coverage area and enter another while a conversation was still in progress. Frequencies would have to be assigned carefully to prevent neighboring transmitters from interfering with one another.

Those complications were substantial.

In return, something extraordinarily valuable became possible: the same limited collection of radio frequencies could be used again and again across the network. The problem was no longer simply how far a radio signal could travel. It was how intelligently the network could divide space.

The Cellular Concept

Once coverage and capacity are treated as different problems, an apparently strange idea begins to make sense. Instead of building one powerful radio station to cover an entire city, divide the city into many smaller coverage areas. Each area has its own radio equipment and serves the subscribers nearby. The transmitter no longer needs to reach everyone. It needs to reach only the devices within its local area. Move far enough away and another radio site takes responsibility.

These individual coverage areas became known as **cells**.

The name is important because it describes more than the shape drawn on a network-planning diagram. A cellular network is fundamentally a way of dividing a large geographic problem into many smaller ones. Rather than treating an entire city as one radio environment, the network creates multiple local radio environments that can be coordinated as parts of a larger system.

This changes the economics of radio spectrum.

Suppose the network has only a limited collection of frequencies available. In a single large coverage area, those frequencies can be used only so many times simultaneously without interfering with one another. Dividing the area into smaller cells creates a new possibility. A frequency being used in one part of the city might also be used in another part, provided the two transmitters are sufficiently separated that their signals do not interfere unacceptably. The

network has not created any new spectrum. It has created more opportunities to use the spectrum it already has. That is the conceptual breakthrough at the heart of cellular communication.

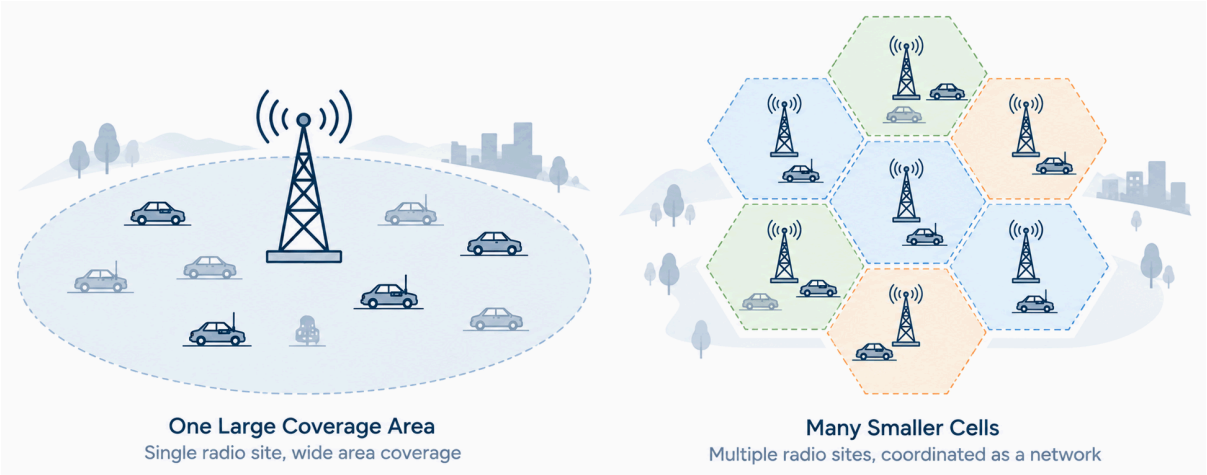


Figure 1.6. *A single large radio coverage area can be divided into many smaller cells, each served by its own radio equipment and coordinated as part of the same network.*

The idea also reverses an instinct inherited from traditional radio systems. A powerful transmitter covering the greatest possible distance is no longer necessarily desirable. Cellular systems benefit from controlling the geographic area in which a transmitter dominates. Smaller coverage areas can make the network more complicated, but they also make repeated use of scarce radio resources possible.

The familiar honeycomb diagrams used to explain cellular networks emerged from the need to reason about this division of space. Hexagons fit together neatly without gaps and provide a convenient mathematical model for neighboring coverage areas. Real radio signals, of course, do not stop at geometric boundaries. Hills, buildings, vegetation, antenna placement, transmitter power, frequency, and countless other factors distort the actual coverage.

A real cell is therefore not a perfect hexagon drawn on the landscape.

ⓘ **Did You Know? — Hexagonal Cells**
The hexagonal cell is primarily a planning and teaching abstraction. Real cellular coverage is irregular and overlapping because radio waves respond to terrain, buildings, antennas, interference, and atmospheric conditions rather than geometric boundaries.

The abstraction is useful because it exposes the larger principle. Once an area is divided into cells, radio resources can be planned geographically.

Neighboring cells can use different frequencies to reduce interference, while cells farther apart can potentially use the same frequencies again.

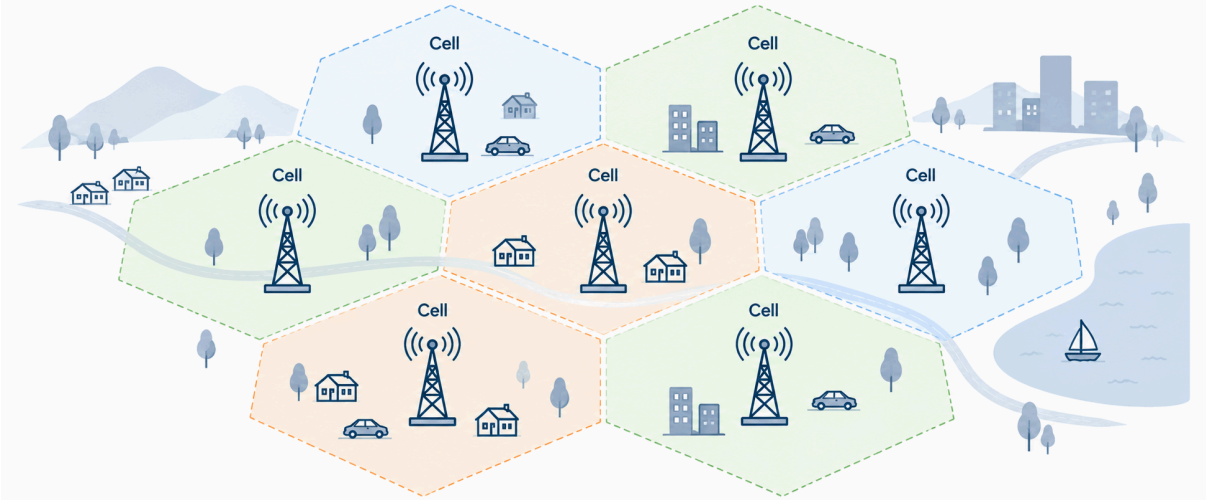


Figure 1.7. *Cells provide a conceptual way to divide geographic coverage so that radio resources can be coordinated locally rather than across one enormous coverage area.*

There is another important consequence.

Capacity can now be increased without necessarily acquiring more spectrum. If demand becomes particularly high in one area, a large cell can potentially be divided into smaller cells, each served by additional radio infrastructure. The same geographic region can then support more independent uses of the available radio resources. Infrastructure has effectively been exchanged for capacity: more base stations, antennas, connections, planning, and coordination allow the network to extract more value from the same spectrum.

This trade-off would become increasingly important as mobile telephony moved from a specialized service toward a mass-market technology. Dense urban areas would eventually contain large numbers of cells of different sizes, while sparsely populated regions could continue using larger cells where coverage mattered more than capacity.

The principle survives even in modern networks. The radio technologies have changed dramatically, and today's systems manage spectrum far more dynamically than early cellular designs, but network operators still add radio sites and reduce effective coverage areas when they need greater capacity in a particular location.

The cellular concept therefore solved one problem by deliberately creating several others.

A collection of independent radio sites was not enough. The sites had to be connected. The network needed to know where subscribers were located well enough to route calls toward them. A subscriber moving through the network could leave the coverage of one site and enter another. Radio resources had to be coordinated so that nearby cells did not interfere excessively.

Most importantly, the limited frequencies available to the network had to be distributed among those cells in a way that allowed them to be used repeatedly. Dividing the landscape created the opportunity. **Frequency reuse** was what made that opportunity valuable.

Frequency Reuse

Dividing a large coverage area into smaller cells created an opportunity, but simply drawing boundaries on a map did not increase capacity. The real advantage appeared when the same radio frequencies could be used more than once.

Imagine that a network has a limited collection of radio channels available. If every transmitter across an entire city used the same channel at the same time, receivers would hear competing signals and communication could become unreliable. One solution would be to assign every cell its own unique frequencies, but that would quickly exhaust the available spectrum. The network would gain smaller coverage areas without gaining much additional capacity.

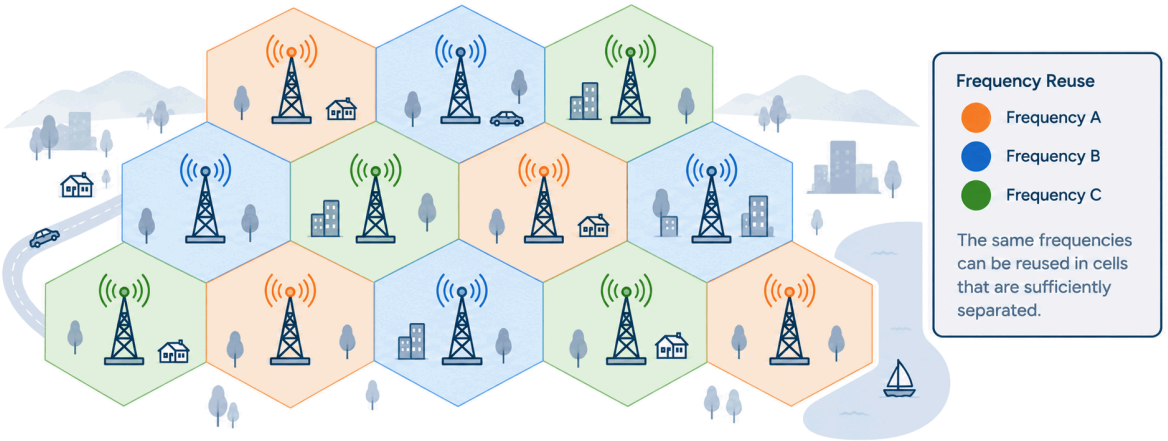


Figure 1.8. *The same radio frequencies can be reused in cells that are sufficiently separated, allowing a limited amount of spectrum to support multiple simultaneous conversations across the network.*

The alternative was to reuse frequencies.

Two cells separated by sufficient distance could use the same radio channel simultaneously because the signal from one cell would be weak enough by the

time it reached the other. The same piece of spectrum could therefore support one conversation in one part of the network and another conversation somewhere else.

This is where the cellular architecture begins to reveal its real power.

The available spectrum has not increased. The network is simply arranging its use geographically. Instead of treating a frequency as a resource that can be used only once across an entire city, it becomes a resource that can be used repeatedly wherever sufficient separation keeps interference within acceptable limits.

Consider a simplified network with several groups of radio channels. One group might be assigned to one cell, another group to its neighbor, and a third group to the next. Move far enough across the network and the first group can be assigned again. The pattern can then repeat.

Early cellular planning often represented this using clusters of idealized hexagonal cells. Each cell within a cluster received a different group of frequencies, while the same groups were reused in other clusters some distance away. Engineers could reason about how closely those frequencies could be repeated while maintaining acceptable communication quality.

The distance between cells using the same frequencies matters because radio signals do not stop at the edge of a cell. A transmitter intended to serve one area continues radiating beyond it, becoming progressively weaker with distance. If another transmitter using the same frequency is too close, a receiver may receive significant energy from both.

The unwanted signal becomes **co-channel interference.**

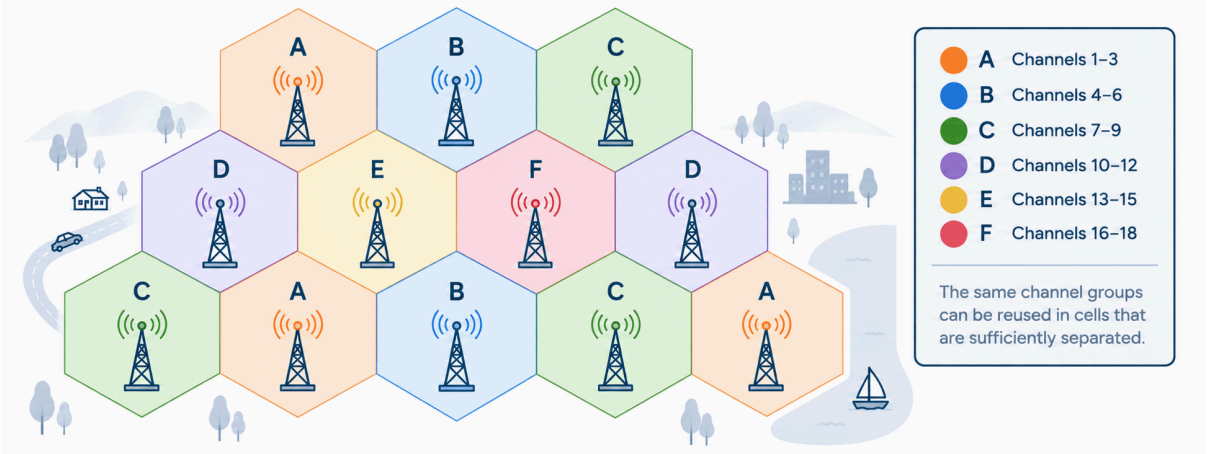


Figure 1.9. *A simplified frequency reuse pattern assigns different channel groups to neighboring cells while allowing those groups to be reused in sufficiently distant cells.*

This means frequency reuse is not simply a matter of repeating the same pattern as often as possible. Reuse that is too aggressive increases interference. Reuse that is too conservative wastes opportunities to increase capacity. Engineers must balance the two. That balance exposes another recurring characteristic of mobile-network design: there is rarely a single quantity that can be maximized without affecting something else.

🔥 **Engineering Insight — Frequency**
Frequency reuse turns geography into a network resource. Capacity depends not only on how much spectrum an operator has, but on how often that spectrum can be reused without creating unacceptable interference.

The consequences are substantial because a set of radio channels capable of supporting a certain number of simultaneous conversations in one cell can support additional conversations when those channels are reused elsewhere. Repeat the process across a large network and a relatively small amount of spectrum can support far more simultaneous communication than would be possible with one city-wide transmitter.

Adding infrastructure can increase that effect further. Where demand becomes high, cells can be made smaller and radio sites placed closer together. Smaller cells reduce the geographic area served by each site and create more opportunities for radio resources to be reused. This does not make capacity unlimited.

More cells require more infrastructure, more coordination, and increasingly careful management of interference. Buildings, terrain, subscriber distribution, antenna design, and traffic patterns complicate the clean geometry of planning diagrams. Real networks are never the perfectly repeating honeycombs shown in simplified illustrations.

Despite that complexity, the underlying principle remains remarkably powerful. Modern LTE and 5G networks manage radio resources very differently from early analog cellular systems: they can schedule users dynamically across time and frequency, adapt transmissions to changing radio conditions, and coordinate interference using techniques unavailable to the engineers who designed the first cellular networks. The simple fixed reuse patterns of early systems are no longer a complete description of how spectrum is managed.

The fundamental insight, however, survives.

> **A scarce radio resource becomes far more valuable when it can be used repeatedly across space.**

That insight solved much of the capacity problem, but it introduced another challenge that ordinary radio systems did not have to solve.

A subscriber was no longer communicating through one transmitter covering an entire region. The network now consisted of many cells, and a moving vehicle could travel from one to another while a conversation was still taking place. The network had learned how to reuse its frequencies—now it had to learn how to follow its subscribers.

Mobility and Handover

Frequency reuse made it possible to serve far more subscribers with the same limited spectrum, but it changed the nature of the network in a fundamental way. A mobile subscriber could now move beyond the area served by one radio site and enter the area served by another.

For someone carrying a modern phone, this happens invisibly. A conversation continues while walking through a city. Music keeps streaming along a highway. A passenger on a train can remain connected while passing through the coverage areas of many different base stations. The user does not choose which radio site should provide service or decide when another would be better.

The network handles the movement.

Early cellular designers faced the same underlying problem with much simpler technology. If a vehicle was communicating through one cell and drove toward the edge of its coverage, the radio signal would gradually weaken. At the same time, the signal from a neighboring cell might become stronger.

Ending the call and asking the subscriber to reconnect through the new cell would hardly provide the experience expected from a telephone network. If cellular telephony was going to support genuine mobility, the active conversation needed to move from one radio channel and base station to another without requiring the subscriber to start again.

This transfer became known as a **handover**, or **handoff** in some terminology. The word sounds simple. The operation is not.

A network first needs some indication that the existing radio connection is becoming unsuitable. It must determine whether another cell can provide a better connection. Radio resources must be available in that cell, and the network must arrange for the conversation to use them. The connection through the fixed network must then be redirected toward the new radio path.

All of this has to happen quickly enough that the subscriber experiences a continuous telephone call.

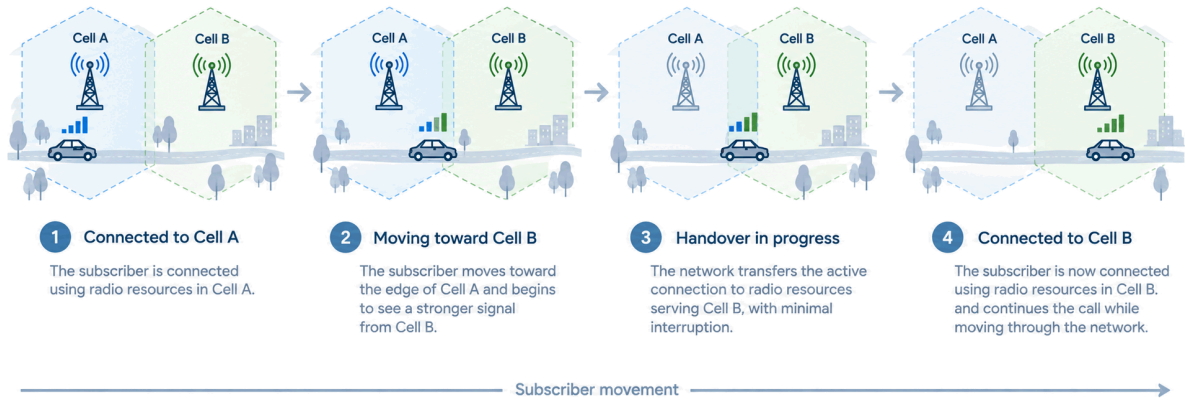


Figure 1.10. *As a subscriber moves away from one cell and toward another, the network must transfer the active connection to radio resources serving the new area.*

The challenge is not merely detecting that another transmitter can be heard. Radio conditions fluctuate constantly. A signal can weaken because a vehicle passes behind a building, travels through a cutting, or encounters interference. Immediately transferring a call whenever another signal briefly becomes stronger could cause the connection to bounce unnecessarily between cells.

Mobility therefore requires decisions.

The network must determine not only *whether* a subscriber can communicate with another cell, but *when* moving the connection becomes preferable to keeping it where it is.

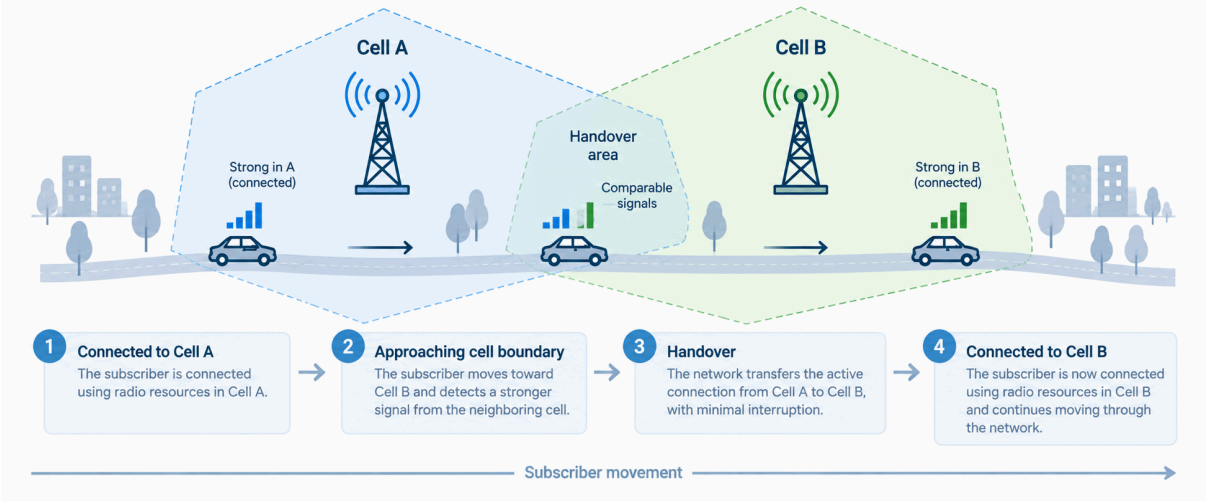


Figure 1.11. During *handover, an active connection is transferred from the radio resources of one cell to those of another as the subscriber moves through the network.*

This is an important departure from the earlier radio telephone model.

With one powerful transmitter covering a large region, the radio connection was conceptually straightforward. The subscriber remained within range of the same radio site until they left its coverage entirely. A cellular network deliberately introduced many smaller coverage areas to gain capacity, which meant the network also inherited responsibility for maintaining service as subscribers crossed between them.

Mobility had become a network problem.

That principle reaches much further than an individual telephone call. A mobile network must deal with subscribers who move while actively communicating, but it must also deal with subscribers who move while doing nothing at all. If someone drives across a country with a phone switched on but never makes a call, the network still needs enough information about their location to find them when an incoming call arrives.

🔥 **Engineering Insight — Mobility**
Mobility is a balance between knowing where a subscriber is and minimizing the work required to maintain that knowledge. Too little location information makes subscribers difficult to find; too much creates unnecessary signaling and consumes network resources.

Tracking every subscriber continuously and precisely would create enormous signaling overhead. Knowing nothing about their location would force the network to search everywhere.

The practical answer lies somewhere between those extremes.

Later mobile systems would develop increasingly sophisticated mechanisms for location areas, registration, paging, cell reselection, measurement reporting, tracking areas, and handover. The terminology and responsibilities would change from one generation to another, but the underlying questions would remain remarkably consistent.

Where is the subscriber?

Which radio connection should serve them?

How much does the network need to know while they are idle?

How can an active connection survive when the answer changes?

For the first cellular telephone systems, solving these problems required much more than radio engineering. Base stations needed coordination. Switching

systems needed awareness of the cellular network. Calls had to be redirected between radio sites. Subscriber location became information the network itself had to manage.

The telephone was becoming mobile, but something more significant was happening underneath it—the network was learning to move with the subscriber.

The Transition Toward Commercial 1G

By the time the cellular concept had taken shape, many of the fundamental problems of mobile telephony were understood. Radio could connect a moving subscriber to the telephone network. Dividing a large service area into cells could increase capacity. Frequencies could be reused when sufficient geographic separation controlled interference. Calls could be transferred between cells as subscribers moved.

Understanding those principles, however, was very different from building a network that ordinary customers could actually use.

A commercial cellular system needed all of these ideas to work together. Radio sites had to be deployed across large geographic areas. Frequencies had to be planned. Switching equipment had to establish calls and coordinate movement between cells. The network needed to identify subscribers, locate them well enough to deliver incoming calls, and connect them to the existing public telephone network.

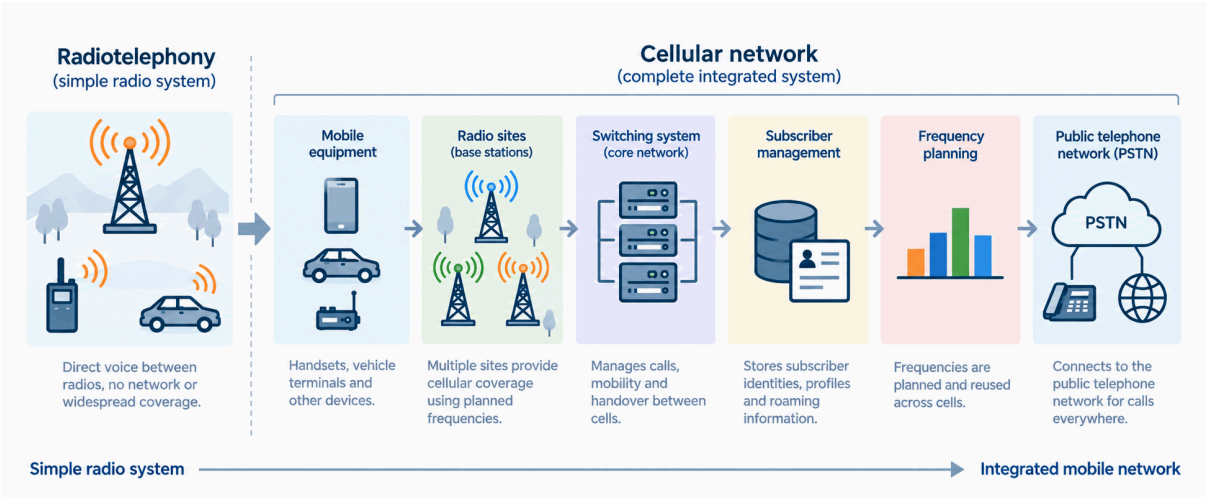


Figure 1.12. *The transition from radio telephony to cellular service required radio sites, switching systems, subscriber management, frequency planning, mobile equipment, and the public telephone network to operate as one coordinated system.*

The mobile equipment also had to become practical enough to sell.

None of these problems existed in isolation. A handset could not be useful without a network. A network could not attract subscribers without suitable devices. Operators could not justify widespread infrastructure without confidence that enough people would use the service. Manufacturers, telecommunications operators, regulators, and governments therefore became participants in the same emerging ecosystem.

Different countries and regions approached the challenge in different ways.

Spectrum allocations differed. Existing telecommunications infrastructure differed. Regulatory environments differed. Equipment manufacturers had their own technologies and commercial interests. As a result, the first generation of cellular networks did not emerge as one universal global system.

Instead, several analog cellular families appeared. In North America, the **Advanced Mobile Phone System**, better known as **AMPS**, became one of the defining first-generation systems. In the Nordic countries, **Nordic Mobile Telephone**, or **NMT**, provided another approach. Other systems appeared elsewhere, each solving broadly similar problems while making different technical and regulatory choices.

These networks would later be grouped under the label **1G**—the first generation of cellular mobile telephony. The term can be slightly misleading when looking backward. Engineers building these systems were not necessarily thinking of themselves as creating "1G" in anticipation of 2G, 3G, 4G, and 5G. The generational language became useful as mobile technology evolved and fundamentally different systems followed.

What distinguished these early cellular networks was not the number attached to them, but the combination of cellular architecture with analog radio technology.

Voice remained analog over the radio interface. The network divided geographic areas into cells, allocated radio channels, established calls, and transferred active conversations as subscribers moved. The cellular principles were modern enough to establish the basic shape of mobile networking, even though the radio technology belonged to an analog era.

✍ **Historical Note — Nordic Countries**
The Nordic countries played an important role in demonstrating what cellular mobile telephony could become. NMT was designed through cooperation between the Nordic telecommunications administrations and supported roaming between participating countries. That ability to cross a national border while retaining mobile service foreshadowed one of the defining ambitions of later European mobile standards.

Commercial deployment also transformed the meaning of mobility.

Earlier mobile telephone systems had demonstrated that a telephone could operate from a vehicle. Cellular networks offered a path toward something much larger: a service capable of supporting growing numbers of subscribers across extensive geographic areas.

At first, the equipment was still expensive and physically substantial. Many early cellular telephones remained vehicle-mounted, while portable units could be heavy and offered limited battery life. Coverage was far from universal, and service was priced well beyond what many consumers could justify. But those limitations were no longer fundamental architectural barriers.

Networks could add cells. Manufacturers could reduce the size and cost of electronics. Operators could expand coverage. Subscriber numbers could grow without requiring a unique radio frequency for every customer. The system had a mechanism through which it could scale.

The achievement should not be underestimated.

For the first time, the telephone number was beginning to separate from a fixed physical location. A person could leave the home or office, enter a vehicle, travel through a network, cross between cells, and remain reachable through the telephone system. Yet the first commercial cellular networks also carried the limitations of the technologies from which they were built.

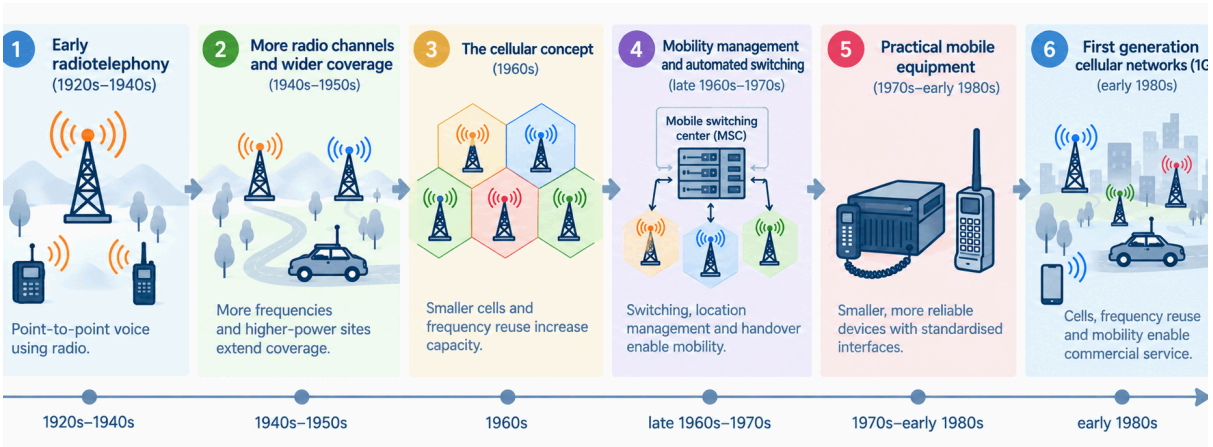


Figure 1.13. *Early radio telephone systems evolved toward commercial cellular networks as cells, frequency reuse, mobility management, switching, and increasingly practical subscriber equipment were brought together.*

Analog radio consumed valuable spectrum. Capacity remained precious. Different regional systems were not automatically compatible. Security was limited. Equipment remained costly and cumbersome. International mobility

depended on cooperation between networks rather than being an inherent global property of the technology.

Cellular architecture had solved the problem of how a mobile telephone network could scale geographically. It had not solved every problem of making mobile communication universal. The first generation had made cellular telephony real. The challenge was now to make it better.

Looking Ahead

The arrival of commercial cellular networks marked a turning point in communications.

For decades, mobility had been constrained by the physical realities of radio. Equipment was large, spectrum was scarce, only a limited number of conversations could take place at once, and providing coverage across a large area often meant allowing a small number of powerful transmitters to dominate that area.

The cellular concept changed the equation.

Instead of trying to make one transmitter reach farther, engineers divided geography into smaller coverage areas. Instead of treating a radio frequency as something that could be used only once across an entire region, they found ways to reuse it at sufficient geographic separation. Instead of assuming that a subscriber would remain connected to one radio site, the network learned to transfer responsibility as that subscriber moved.

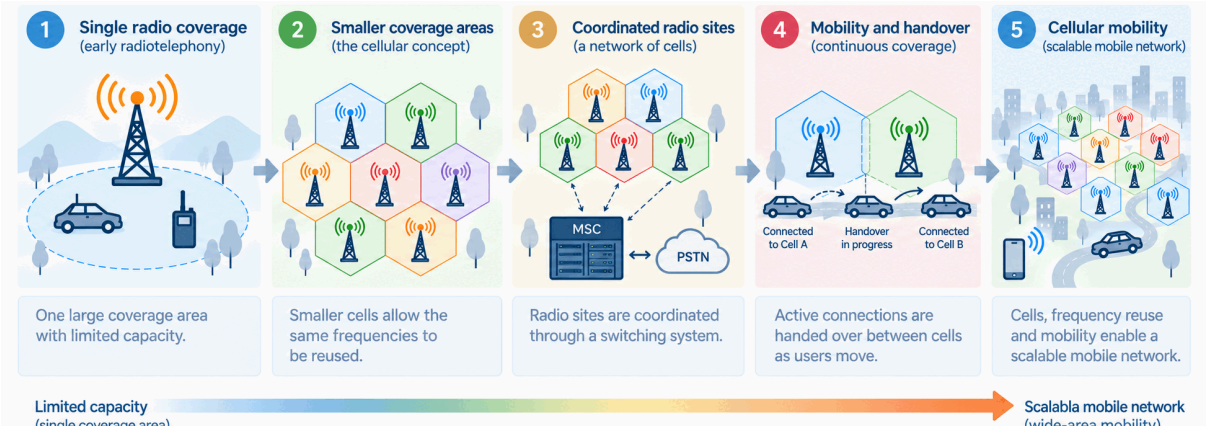


Figure 1.14. *The cellular concept brought together smaller coverage areas, frequency reuse, coordinated radio sites, and mobility to transform limited radio spectrum into a scalable mobile telephone network.*

These ideas established a foundation that would survive every generation that followed.

The details would change enormously. Analog voice would give way to digital communication. Mobile networks would begin carrying messages and data. Circuit-switched connections would eventually coexist with, and then largely give way to, packet-based communication.

SIM cards, encryption, authentication, roaming agreements, Internet connectivity, multimedia services, smartphones, and entirely new radio technologies would become part of the system.

But the underlying engineering problems would remain surprisingly familiar. Spectrum would still be finite. Coverage and capacity would still compete for engineering attention. Radio resources would still need to be shared among many subscribers. Networks would still need to know enough about the location of a device to provide service. Connections would still need to survive movement. Infrastructure built by different companies, operators, and countries would still need ways to work together.

What changed was the scale.

The earliest mobile telephone systems served relatively small numbers of subscribers using expensive equipment. Cellular architecture created a path toward networks capable of serving millions, and eventually billions, of devices. Once that path existed, the question was no longer whether mobile telephony could work.

The question became how far it could go.

The first generation of commercial cellular networks provided the first real answer. They took the principles developed through decades of radio engineering and turned them into operational systems that people could actually use. Their radio technology was analog, their devices were cumbersome by modern standards, and their capabilities were modest.

Yet they established something extraordinary: a telephone network in which the endpoint could move. That was enough to begin a revolution. The age of **1G** had arrived.

1G: The Analog Generation

How do you turn the elegant idea of a cellular network into something that millions of people can actually use?

Dividing a geographic area into cells provides the beginning of an answer. Frequencies can be reused, radio coverage can be expanded by deploying additional sites, and an active connection can be transferred as a subscriber moves from one cell to another. On paper, these principles provide a compelling solution to the limitations of earlier mobile radio systems.

A commercial mobile telephone network, however, cannot exist as a collection of cells on a map.

Someone sitting in a car must be able to pick up a telephone, request a connection, and speak with another person through the ordinary telephone network. Someone calling that mobile subscriber must somehow reach them even though the network cannot know exactly where they are. If the vehicle moves while the conversation is taking place, the radio connection must survive the journey between cells.

Thousands of other subscribers may be attempting to do the same thing, all competing for a limited collection of radio channels. The network must make all of this appear ordinary.

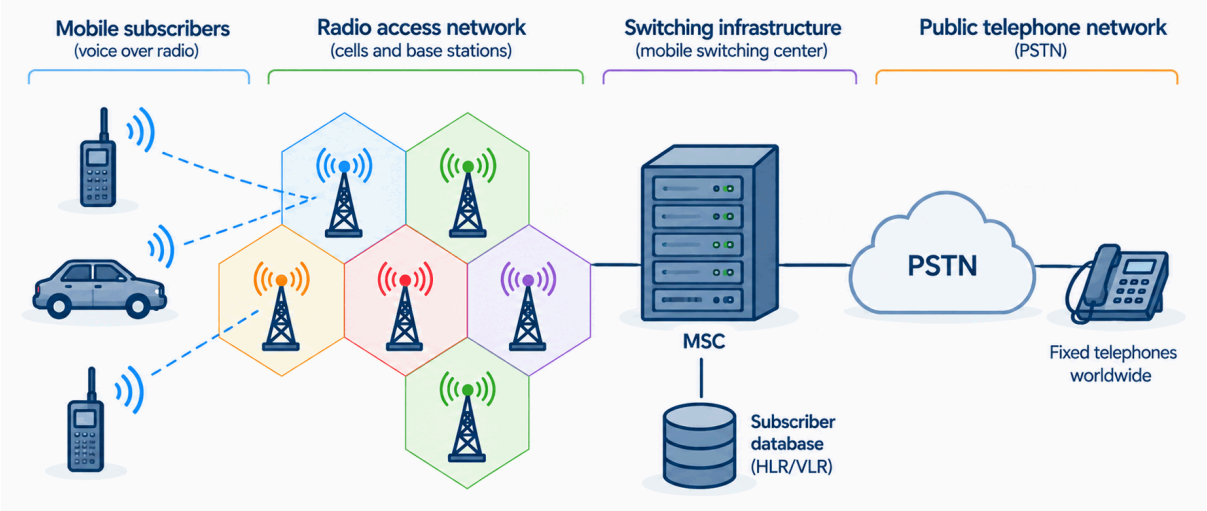


Figure 2.1. *A first-generation cellular network connected mobile subscribers through radio base stations and switching infrastructure into the existing public telephone network.*

That was the challenge faced by the first generation of commercial cellular systems. The cellular concept had provided a way to organize radio coverage and reuse scarce spectrum, but a practical service required an entire system

around it: radio equipment, base stations, switching infrastructure, subscriber information, channel allocation, call control, mobility management, and connections into the existing telephone network.

From the subscriber's perspective, much of this complexity disappeared behind a familiar interface. There was a telephone handset, a number to dial, and a conversation. The remarkable part was that the wire between the subscriber and the telephone exchange had been replaced by a radio connection that could change as the subscriber moved.

That radio connection was still fundamentally analog.

Speech entering the handset was converted into an electrical signal whose continuously varying characteristics represented the sound of the speaker's voice. The radio system carried that information across an allocated channel to the network, where it could ultimately become part of an ordinary telephone connection. In the other direction, the voice of the distant caller traveled through the network and across the radio channel to the mobile subscriber.

The cellular architecture surrounding that conversation was therefore solving a much broader problem than simply transmitting speech.

It had to decide which radio channel the subscriber should use. It had to establish a path through the network when a call began and release those resources when the call ended. It had to determine when a moving subscriber should be transferred to another cell. If the network extended across regions or national borders, it also needed mechanisms through which subscribers could receive service away from their usual location.

Different parts of the world developed different systems to accomplish these tasks. NMT emerged from Nordic cooperation, while AMPS became a major cellular system in North America, alongside other first-generation technologies developed and deployed elsewhere. They were not identical implementations of a single worldwide cellular standard. They reflected different spectrum allocations, regulatory environments, engineering decisions, and commercial circumstances.

Yet beneath those differences, the systems were attempting to solve many of the same fundamental problems. They had to connect moving subscribers, share finite spectrum, coordinate cells, establish calls, and integrate mobile radio with a telephone network that had been built around fixed endpoints.

This is what makes 1G important beyond its place in a historical timeline. It transformed cellular engineering from a concept into an operating network and, in doing so, established responsibilities that every later generation would inherit in one form or another. The technologies used to fulfill those

responsibilities would change dramatically, but the network would still need to allocate radio resources, establish communication, manage mobility, identify subscribers, and connect them to services elsewhere. That emphasis on enduring engineering principles rather than treating generations as collections of acronyms is central to understanding how mobile networks evolved.

To understand what those first networks were actually doing, the best place to begin is not with their acronyms or switching equipment, but with the thing the subscriber experienced directly: a voice traveling across an analog cellular radio channel.

Analog Cellular Voice

> **What actually traveled through the air when someone made a call on a first-generation cellular network?**

The obvious answer is *voice*, but that hides an important part of the engineering. The network did not understand the words being spoken, nor did it need to. Its job was to preserve the continuously changing electrical signal produced by the speaker's voice well enough for that sound to be reconstructed at the other end of the connection.

A microphone inside the telephone converted variations in air pressure into a corresponding electrical signal. That signal changed continuously with the sound of the speaker's voice. Rather than first representing those changes as a sequence of digital numbers, as later mobile systems would, a 1G radio carried the voice using an analog radio transmission.

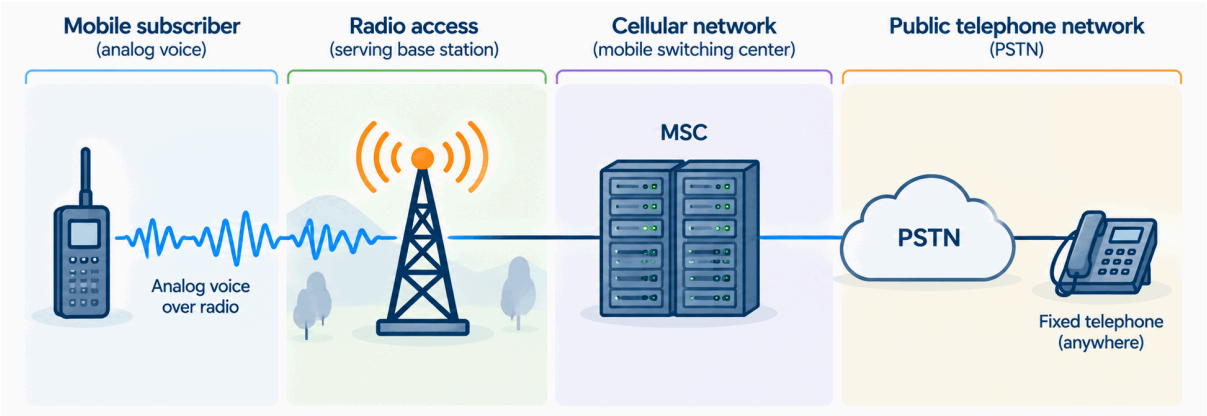


Figure 2.2. *During a 1G call, the mobile telephone used an analog radio connection to reach a serving base station, while the cellular network connected the conversation into the wider telephone network.*

The distinction matters because the radio channel was occupied for the duration of the conversation. Once a call had been established, radio resources were assigned to carry the connection between the mobile

telephone and the serving base station. Those resources remained associated with the call while the subscriber spoke, listened, or simply remained silent. The telephone network beyond the radio system similarly maintained a connection through which the conversation could travel.

This made the mobile telephone experience conceptually similar to the fixed telephone service people already understood. A subscriber requested a call, a connection was established, the conversation occupied resources while it remained active, and those resources were released when the call ended.

The important difference was the first part of that path.

A fixed telephone had a physical connection leading toward the telephone exchange. The mobile subscriber had a radio channel shared within a cellular system. That channel had to come from the finite collection of frequencies available to the cell, which meant every active conversation consumed part of the cell's available radio capacity.

The relationship between voice and capacity was therefore unusually direct.

A cell with all of its usable voice channels occupied could not accommodate another call merely because the new subscriber had an excellent signal. The network first needed a suitable channel to become available. Good coverage meant that the subscriber could communicate with the base station; it did not guarantee that the base station had radio resources available for another conversation.

This is the same distinction between coverage and capacity that made the cellular concept necessary in the first place, now experienced by an individual caller. There was also more happening over the radio interface than the conversation itself. A functioning cellular system needed ways to establish calls, identify the appropriate radio resources, alert a subscriber to an incoming call, supervise connections, and coordinate movement between cells. The voice channel was therefore only one part of a larger control system that made mobile telephony possible.

That separation is easy to miss because the subscriber experienced only the result. Pick up the handset, enter a number, and wait for the other person to answer. Behind those familiar actions, the network had to locate resources and create a path spanning two very different environments: a radio connection whose characteristics changed as the subscriber moved and a telephone network built to maintain calls between endpoints.

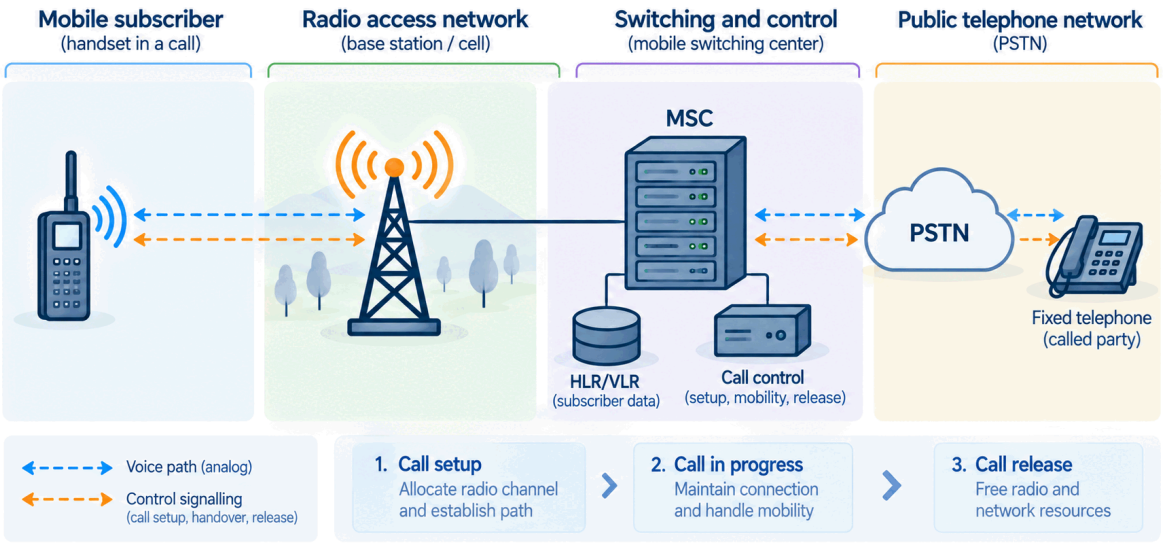


Figure 2.3. *A cellular call depended on both the radio resources carrying the conversation and control functions used to establish, manage, and eventually release the connection.*

The analog nature of the radio link also had consequences that would become increasingly important as cellular service grew. The network could manage access to a radio channel, but an analog transmission traveling through the air was fundamentally different from a private wire entering a home. Radio signals could extend beyond the intended subscriber, and anyone with suitable receiving equipment could potentially observe transmissions within range.

Later generations would approach the radio connection very differently, particularly as digital representation made sophisticated authentication, encryption, and error correction practical while enabling radio-access techniques that could use the available spectrum more efficiently. For 1G, however, the immediate achievement was more fundamental: analog radio had been incorporated into a cellular architecture capable of providing recognizable telephone service to a moving subscriber.

Making one such call was an engineering accomplishment.

Making the same service work across cities, regions, and eventually national borders required something larger. Operators needed complete cellular systems, and different parts of the world did not all choose the same one.

NMT, AMPS, and Other 1G Systems

> **If the engineering problems were broadly the same, why did the world end up with several different first-generation cellular systems?**

There was no single global organization defining how commercial cellular telephony should work. Telecommunications had developed largely within national and regional boundaries, spectrum was allocated differently, regulatory structures varied, and operators had inherited their own infrastructure and technical traditions. When cellular networks began moving from engineering concept to commercial reality, those differences naturally influenced what was built.

The result was not one 1G network, but a collection of analog cellular systems. Among the most significant were **Nordic Mobile Telephone (NMT)** and the **Advanced Mobile Phone System (AMPS)**. Other analog systems appeared in different markets, each providing its own answer to the same basic challenge: how to turn cells, radio channels, switching, and mobility into a usable telephone service. The details differed, but the responsibilities of the network were strikingly similar.

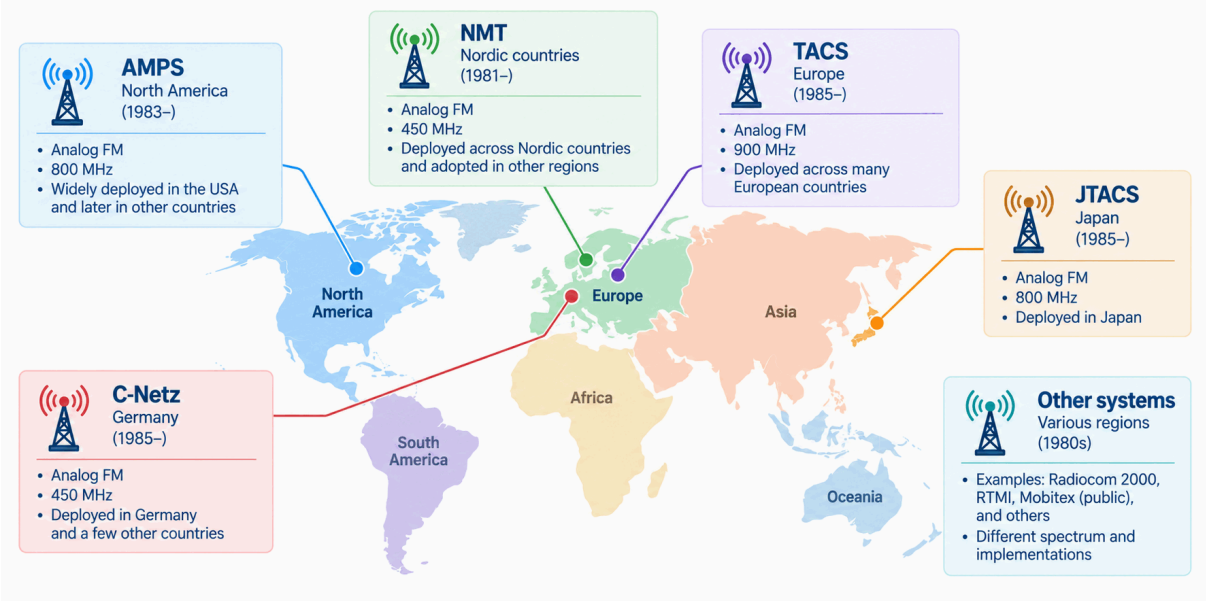


Figure 2.4. *First-generation cellular networks developed as several regional system families, including NMT and AMPS, rather than as a single worldwide mobile standard.*

NMT emerged from cooperation between the Nordic telecommunications administrations. That cooperation mattered because the ambition extended beyond making cellular telephony work within a single country. A subscriber traveling between participating Nordic countries should be able to continue using mobile service rather than discovering that the telephone became useless at the border.

This required agreement.

Networks needed compatible radio technology and common expectations about how mobile equipment would interact with the infrastructure. Operators needed procedures through which subscribers belonging to one network could receive service from another. Manufacturers needed sufficiently clear technical requirements to build equipment capable of operating within the system.

The significance of that approach reaches far beyond analog telephony. Mobile communication becomes dramatically more useful when the network is treated as an interoperable ecosystem rather than an isolated collection of equipment.

AMPS approached the emerging cellular market from a different regional environment. Developed for the North American market, it became another important example of how the cellular principles could be transformed into a large commercial system. Like NMT, it combined cells, analog radio channels, switching, call control, and mobility into an architecture capable of supporting subscribers across an expanding geographic footprint.

The two systems did not need to be identical to demonstrate the same underlying idea. A subscriber moving through an NMT network and one moving through an AMPS network presented their respective systems with many of the same engineering problems. Both needed radio coverage. Both had to allocate channels from finite spectrum. Both needed to establish and terminate calls. Both needed to know enough about mobile subscribers to deliver service, and both needed to maintain active conversations as those subscribers moved between cells.

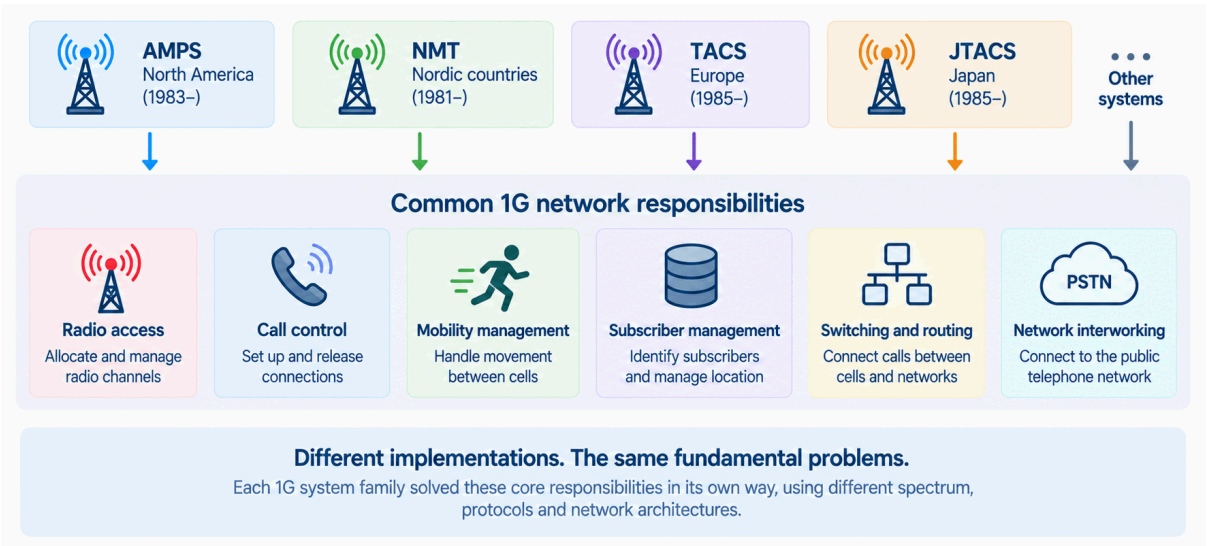


Figure 2.5. *Different 1G system families could use different implementations while still solving the same fundamental problems of radio access, switching, mobility, and connection to the telephone network.*

Their implementations reflected the environments in which they had been created.

This is an important distinction when thinking about mobile-network generations. A generation is not necessarily a single technology deployed identically everywhere. Even at the beginning of commercial cellular networking, regional systems could share broad architectural principles while differing in frequencies, signaling, equipment, network design, and operational details.

Those differences had practical consequences.

A telephone designed for one cellular family could not simply be assumed to operate on another. The radio frequencies might differ, but incompatibility extended beyond the tuning of the transmitter and receiver. The device and network had to agree about how calls were requested, how channels were assigned, how subscribers were recognized, and how mobility was managed.

Crossing from one technological ecosystem into another could therefore mean losing service entirely.

Within a compatible ecosystem, however, something much more interesting became possible. The subscriber could begin to separate not only from a fixed telephone line, but also from a single physical network location. NMT's regional cooperation demonstrated the importance of this idea particularly clearly: mobility could extend beyond the coverage of one operator or one country when networks agreed on the technical and operational mechanisms required to support it.

That possibility would eventually become one of the defining expectations of mobile communication.

It also revealed an engineering lesson that would grow more important with every generation that followed. A mobile network is not useful merely because its individual components work. Devices, radio infrastructure, switching systems, operators, and eventually networks in different countries must agree on enough common behavior for the whole system to function.

Interoperability was becoming as important as radio engineering.

For the subscriber, however, all of this cooperation remained largely invisible. The telephone still appeared to communicate with whichever radio site happened to serve the local area. Somewhere behind that site, the network had to make decisions about where the call should go, which resources it should use, and how the mobile connection related to the much larger telephone system.

The cell provided the radio connection—something else had to turn thousands of those cells into a network.

Base Stations and Switching

> **A cell can provide the radio connection to a mobile telephone, but where does the call go after the radio signal reaches the antenna?**

The answer reveals an important characteristic of cellular networks that remains true today. The radio network is only one part of the system. Somewhere behind the antennas and transmitters, fixed infrastructure must turn thousands of individual radio connections into telephone calls that can reach other subscribers.

The base station sits at the boundary between those two worlds.

Within its coverage area, a base station provides the radio resources through which mobile telephones communicate with the network. It transmits toward the mobiles, receives their transmissions, and provides access to the channels assigned to the cell. From the perspective of a subscriber, it is the nearest visible part of the cellular infrastructure. But a base station cannot operate as an isolated radio tower.

Suppose a subscriber in one cell wants to call an ordinary telephone across the city. The base station can receive the mobile transmission, but it still needs some way to connect that call into the wider telephone network. If the destination is another mobile subscriber, the network must determine where that subscriber can be reached and arrange a connection through the appropriate radio infrastructure.

The base stations therefore need to be connected to a switching system.

Telephone switching was not new. Fixed telephone networks had long used exchanges to establish connections between subscribers. Cellular telephony added a complication that fixed networks did not normally face: one end of the connection could move.

A fixed telephone number was associated with a line terminating at a known place in the network. A mobile subscriber could be served by one base station now and another somewhere else later. The switching infrastructure therefore had to participate not only in establishing the call, but also in maintaining the relationship between that call and a changing radio connection.

This made the cellular switch more than a gateway into the public telephone network—it became part of the machinery of mobility.

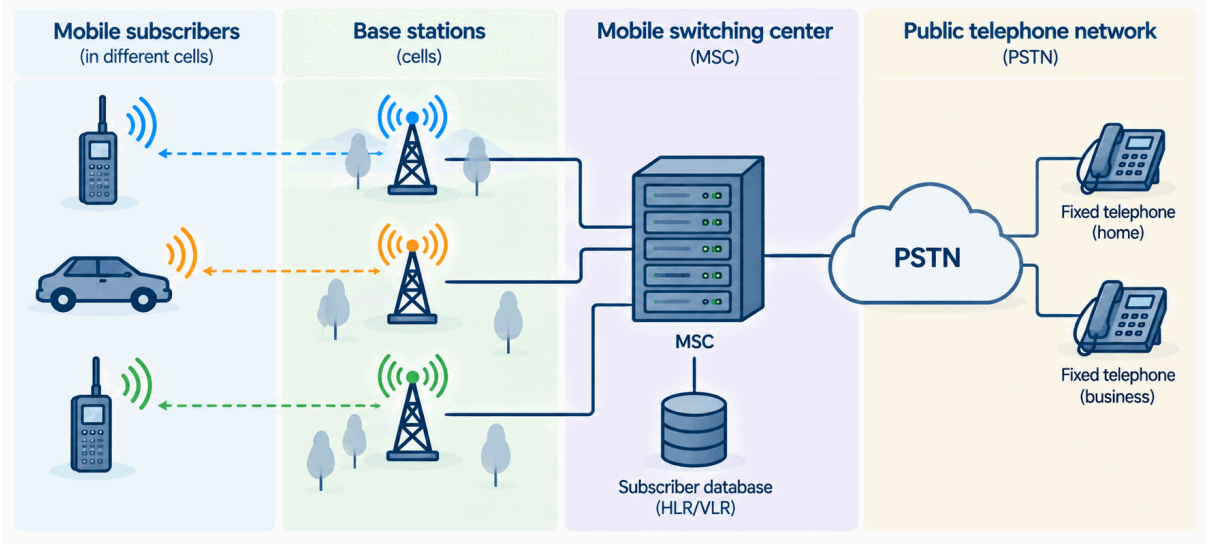


Figure 2.6. *Base stations provide radio access within individual cells, while switching infrastructure connects those cells to one another and to the wider public telephone network.*

Consider a call from a mobile subscriber to a fixed telephone. The mobile first communicates through the base station serving its current cell. The cellular switching system then establishes the appropriate path toward the public telephone network, which carries the call onward toward its destination. In the opposite direction, an incoming call must somehow find the mobile subscriber before the network can determine which radio resources should be used to reach them.

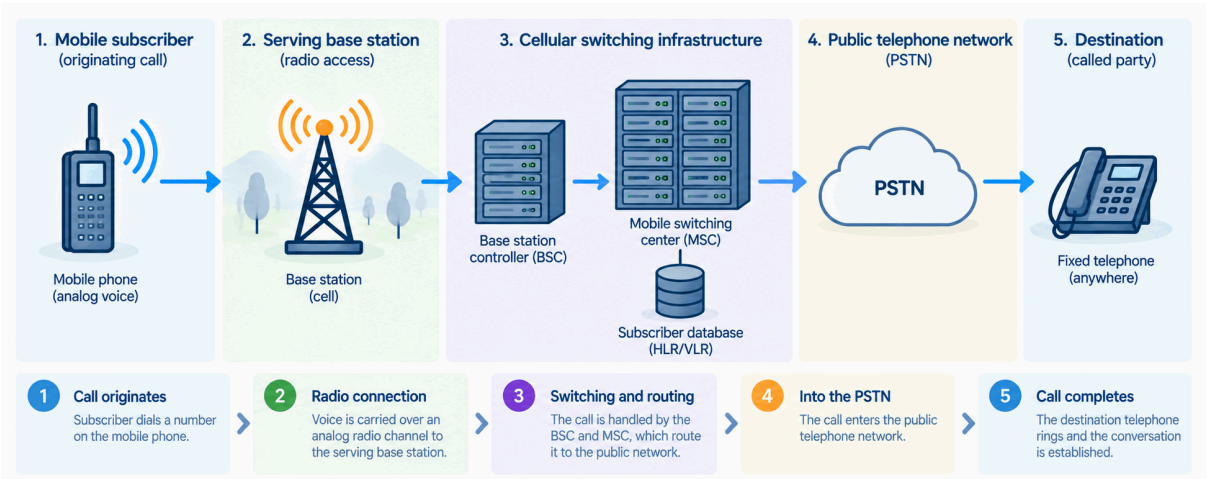


Figure 2.7. *A call from a mobile subscriber passes through the serving base station and cellular switching infrastructure before entering the public telephone network and reaching the destination.*

The same architecture also provides the foundation for calls between mobile subscribers. Even when two telephones happen to be geographically close, their conversation is not simply a direct radio exchange between the devices.

Each mobile communicates with the cellular infrastructure, while the network establishes and manages the connection between them.

That distinction changes how we should think about the antenna towers scattered across a landscape.

They may be the most visible part of a cellular network, but they are not the network itself. The intelligence required to coordinate calls, mobility, and connections to other networks extends beyond the individual cell. Cellular telephony is therefore distributed by its very nature: radio coverage is provided locally, while broader network functions coordinate what happens across those local coverage areas.

The arrangement also allows the radio and switching problems to be treated differently. Additional base stations can be deployed where more coverage or capacity is required, while switching infrastructure can serve many cells and concentrate their traffic into the wider telephone system. Growth in one part of the network does not necessarily require an identical change everywhere else.

For 1G systems, the exact organization and terminology varied between cellular families, but the underlying responsibilities remained recognizable: radio equipment served the cells, switching infrastructure connected calls, and the two worked together to provide mobility. Our central concern is this relationship rather than a catalog of generation-specific network elements.

A call, however, could not begin merely because a base station and switch existed. The radio spectrum available to each cell was still finite, and several subscribers might want to make calls at the same time. Before the switching system could provide the familiar experience of a telephone connection, the radio side of the network had to answer a more immediate question:

Which channel should this subscriber use?

Channel Allocation

A subscriber presses the call button, the base station is within range, and the switching network is ready to establish the connection. One problem remains:

Which radio channel will carry the call?

The answer cannot simply be "any available frequency."

The spectrum assigned to a cellular system has to be organized into usable radio channels, and those channels must be distributed across the network in a way that respects the frequency-reuse plan. A channel that is perfectly suitable in one cell may deliberately be unavailable in a neighboring cell

because using it there could create unacceptable interference. The collection of channels available to a subscriber therefore depends partly on where that subscriber happens to be.

Within a cell, those radio resources must then be shared among everyone attempting to use the service.

Most of the time, a mobile telephone is not carrying a conversation. It may be switched on and reachable, but permanently reserving a voice channel for every subscriber would waste an enormous amount of spectrum. Instead, the network assigns the resources needed for a conversation when the subscriber actually requires them and releases those resources when the call ends.

This allows a relatively small collection of radio channels to serve a much larger population of subscribers.

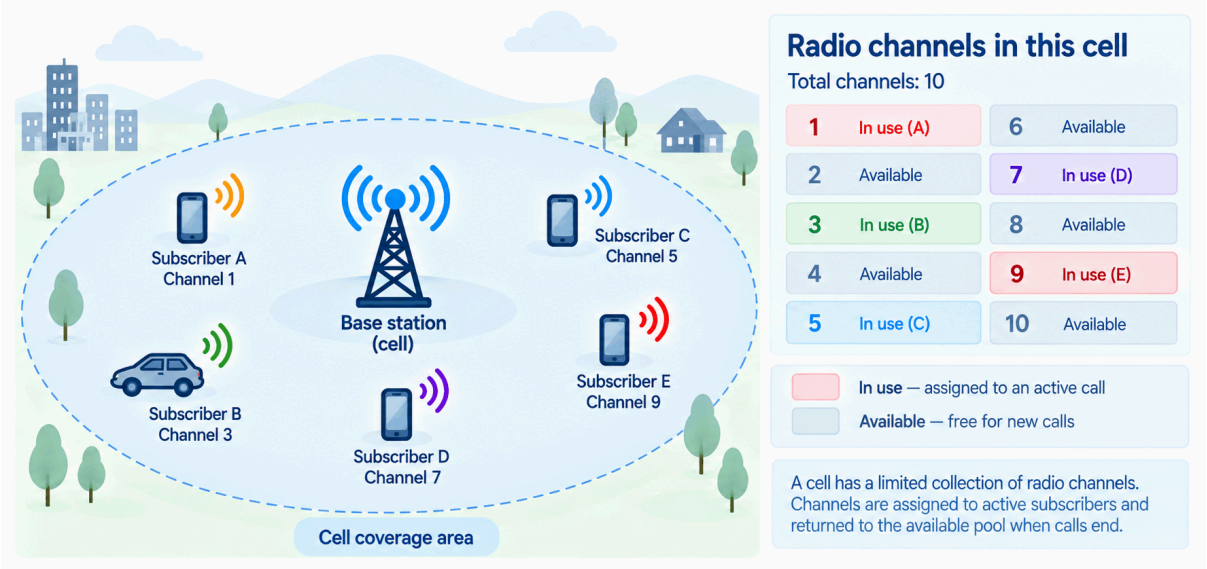


Figure 2.8. *A cell has a limited collection of radio channels available for calls. Channels are assigned to active subscribers as needed and returned to the available pool when those calls end.*

Imagine a cell in which several subscribers are already holding conversations. Each active call occupies radio resources from the collection available there. When another subscriber requests a call, the network looks for a suitable free channel and, if one is available, assigns it to that connection. The subscriber does not need to know which channel has been selected. From their perspective, the telephone simply begins establishing the call. Behind that familiar experience, the network has made a resource-allocation decision.

This is an important change from thinking about radio channels as permanently belonging to particular users. The subscriber has access to the cellular network, not ownership of a particular voice frequency. Radio

resources can therefore be allocated dynamically according to demand. One subscriber may use a channel for several minutes, end the call, and release it. Shortly afterward, another subscriber in the same cell may use that same resource for an entirely unrelated conversation.

The cellular network is continually recycling its scarce radio capacity.



Figure 2.9. *A radio channel is temporarily assigned when a call requires it and returned to the cell's available resources when the connection ends, allowing the same capacity to serve many subscribers over time.*

The arrangement works well while demand remains below the available capacity. Problems appear when too many subscribers want to communicate at once. If every suitable voice channel in a cell is already occupied, another call cannot simply be added. The subscriber may have excellent radio coverage and the switching network may be operating normally, yet the call still cannot proceed until an appropriate radio resource becomes available.

The network is not out of range—it is out of capacity. This is why traffic patterns matter so much in cellular engineering. A network does not need enough radio channels for every subscriber to hold a conversation simultaneously, because that would be extraordinarily inefficient. Instead, capacity can be planned around the expectation that only some portion of the subscriber population will be active at any particular moment.

The difficulty is that subscribers do not distribute themselves evenly.

A business district may experience heavy demand during the working day. A highway can become busy during commuting hours. A stadium or public event can concentrate large numbers of subscribers into a small geographic area. The amount of spectrum has not changed, but the number of people competing for the radio resources of particular cells has.

Cellular capacity is therefore both finite and local. Adding subscribers somewhere else in the network may have little effect on a particular cell, while a sudden concentration of activity within that cell can exhaust its available channels. This is one reason cellular architecture is so powerful: capacity can be increased where demand exists by changing the radio plan, deploying additional infrastructure, or reducing the area served by individual cells so that spectrum can be reused more frequently.

Channel allocation also becomes more interesting when the subscriber is moving. A call may begin using a channel in one cell, but if the vehicle travels into another cell, that original radio resource cannot simply follow it through the landscape. The destination cell must have suitable resources available if the conversation is to continue there.

The network is therefore managing something more subtle than a collection of frequencies. It is managing temporary access to a scarce resource whose availability depends on geography, interference, existing traffic, and movement.

For an individual caller, all of those decisions lead to a much simpler experience: a channel becomes available and the telephone can make a call. But allocating the channel is only one step. The network still has to turn the subscriber's request into a connection, find the destination, alert the other party, and create a path through both the cellular and telephone networks.

A radio channel makes the conversation possible—the network must still establish the call.

Call Setup

> **What actually happens in the few seconds between requesting a call and hearing the other telephone begin to ring?**

To the subscriber, almost nothing appears to happen. A number is entered, the call is requested, and after a brief pause the familiar ringing tone arrives. Yet that pause hides one of the central responsibilities of the cellular network. A telephone conversation cannot begin until the network has transformed the subscriber's intention—*I want to speak to this person*—into an actual connection spanning the radio system and the wider telephone network.

The process begins before a voice channel is carrying any conversation.

A mobile telephone needs a way to communicate its request to the network without already having a voice channel assigned to it. This is why the control functions surrounding the analog radio connection are so important. They

provide the means through which the mobile and network can exchange the information needed to begin establishing a call.

Consider a subscriber sitting in a car who enters the number of a fixed telephone and requests a call. The mobile communicates with the cellular network through the serving cell, identifying the request and the destination being called. The network can then determine whether suitable radio resources are available and begin establishing the connection through its switching infrastructure.

Only when those pieces are in place does the subscriber receive the radio resources needed for the conversation itself.

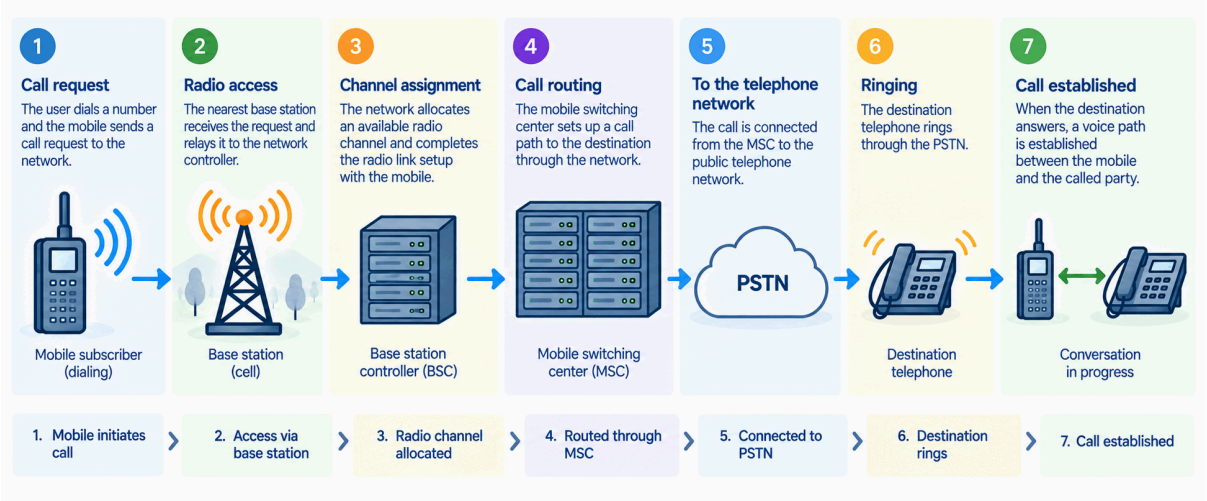


Figure 2.10. *An outgoing 1G call begins with signaling between the mobile and the cellular network. The network assigns suitable radio resources and establishes a path through the switching system toward the destination telephone.*

There is an important separation hidden inside this sequence.

The radio channel carrying the subscriber's voice is not responsible for deciding where the call should go. Nor does the base station simply hear a dialed number and somehow connect itself directly to the destination. The network uses signaling to coordinate the request, switching to establish the path, and radio resources to carry the mobile portion of the resulting conversation.

In other words, **establishing a connection and using a connection are different activities.**

That distinction becomes even clearer when the call travels in the opposite direction. Suppose someone using an ordinary telephone calls a mobile subscriber. The fixed telephone network can deliver the call toward the cellular

system, but the mobile is not attached to a wire terminating at a known socket. It may be parked outside an office, traveling along a highway, or sitting somewhere else within the service area.

Before the call can be delivered, the cellular network has to find it.

The network does not necessarily need to know the subscriber's exact physical position. What it needs is enough knowledge of the subscriber's whereabouts to determine where to search. The appropriate radio infrastructure can then attempt to alert the mobile that an incoming call is waiting. If the telephone responds, the network can assign the necessary radio resources and complete the connection.

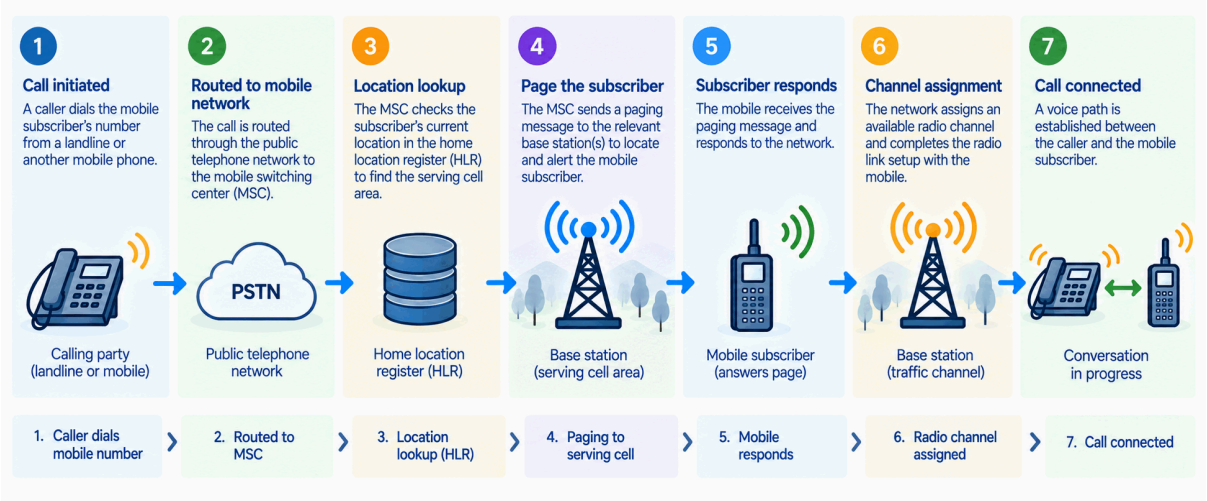


Figure 2.11. *For an incoming call, the cellular network must first locate and alert the mobile subscriber before assigning radio resources and completing the connection.*

This turns an apparently simple telephone call into a coordination problem.

The network has to bring together information about the subscriber, the serving area, available radio resources, and the requested destination. The base station provides access to the mobile, while the switching infrastructure connects that radio access into the larger telephone system. Control signaling coordinates the steps that allow those separate pieces to behave like one service.

The exact procedures differed among first-generation systems, but the engineering requirement was common to all of them: a mobile subscriber could not be treated like a fixed telephone line because the network could not assume that the subscriber remained attached to one permanent physical endpoint.

Once the call is established, much of that machinery disappears from the subscriber's awareness. The allocated radio channel carries the conversation, the switching system maintains the path through the network, and the participants simply talk. But a cellular telephone is still moving. A subscriber who begins a call in one cell may continue driving until the original radio connection is no longer good enough to sustain the conversation. The network has already established a call and committed resources to it, so simply abandoning that connection at the cell boundary would undermine the very mobility that cellular architecture was designed to provide.

Establishing a call is therefore only half of the problem—the network must also keep it alive while the radio world around it changes.

Handover in Analog Networks

A call has been established, a radio channel has been assigned, and the subscriber is talking while driving through the network.

What happens when the car reaches the edge of the cell?

The conversation itself has no awareness of geography. The people speaking may not even realize that the radio conditions are changing. From the network's perspective, however, the connection is becoming increasingly fragile. The mobile is moving farther from the base station serving the call, obstacles may be weakening the signal, and another base station may now provide a better radio path.

Allow the process to continue for too long and the call will eventually be lost.

The obvious solution is to move the radio connection to the neighboring cell. Doing so while preserving an existing telephone call is considerably more difficult than simply establishing a new one. The network already has a conversation in progress, resources have been allocated, and a path through the switching system exists. Changing the radio endpoint must therefore become part of that existing connection rather than replacing it with an entirely new call.

This is the problem that **handover** solves.

The first challenge is recognizing when that transfer should occur. Radio signals rarely change smoothly. A vehicle may pass behind a building, descend into a valley, travel beneath a bridge, or encounter interference from another transmission. Signal conditions can improve and deteriorate even when the subscriber has barely moved. A weak signal at one moment does

not necessarily mean that the call should immediately be transferred elsewhere.

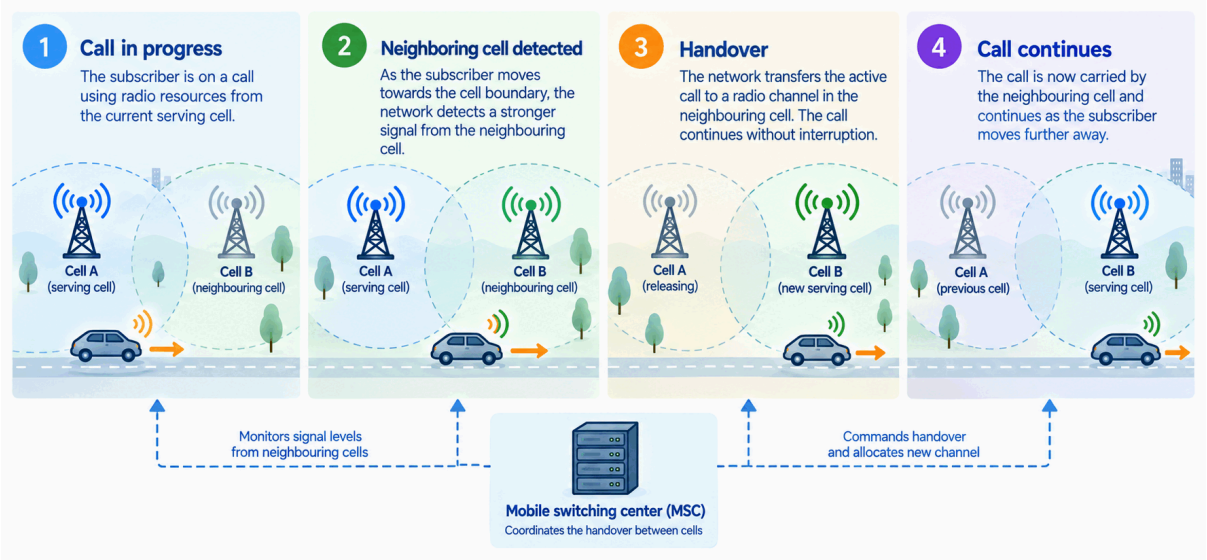


Figure 2.12. *As a subscriber moves away from the serving cell, the network can transfer the active call to radio resources in a neighboring cell so that the conversation can continue.*

The network therefore needs evidence that another radio connection would provide a better way to continue the call.

In an analog cellular system, measurements of the radio connection can help the infrastructure determine that the serving cell is becoming unsuitable and that a neighboring cell may be a better choice. The network can then consider whether the target cell has appropriate radio resources available.

That final condition is easy to overlook. A neighboring cell may provide excellent radio coverage while having no suitable channel available for the incoming call. Handover therefore brings mobility and capacity together. The network is not merely searching for a stronger signal; it is searching for a radio connection capable of taking responsibility for a conversation that already exists.

Once a suitable target has been identified, the network must coordinate the transition. A channel is made available in the new cell, the mobile is directed toward the new radio resource, and the switching infrastructure changes the path carrying the call so that the conversation continues through the new base station.

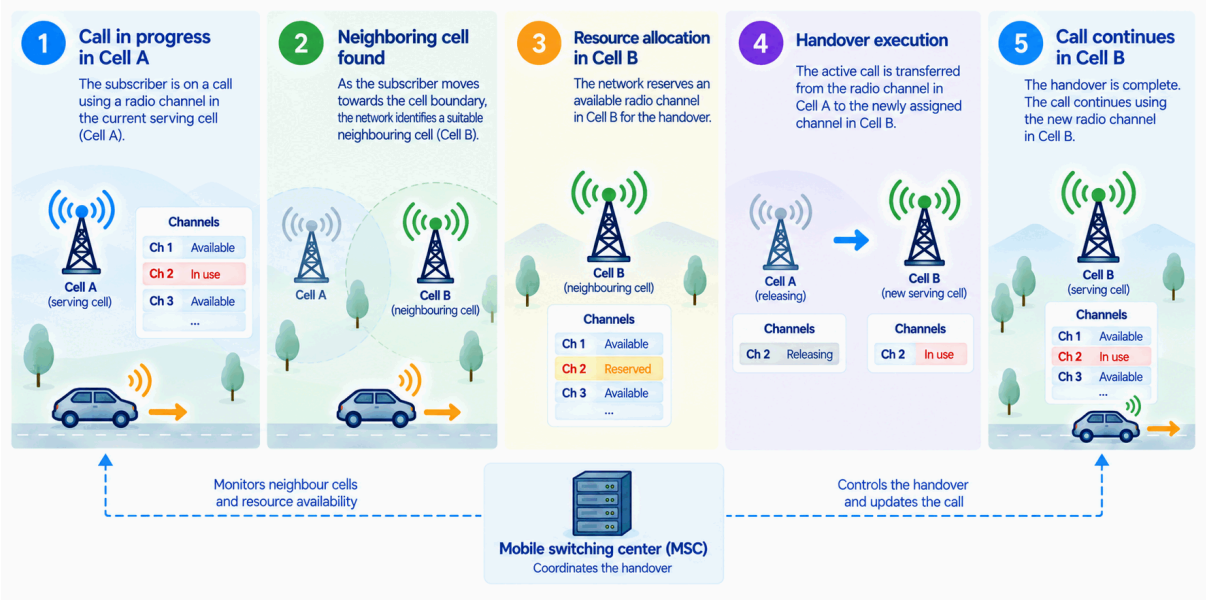


Figure 2.13. *A handover requires both a suitable neighboring cell and radio resources within that cell. The active call is transferred from the original radio connection to the newly assigned one.*

For a brief moment, a considerable amount of network activity is taking place simply to preserve something the subscriber already has. That is one of the more revealing characteristics of mobility. Establishing a connection is not enough. The network must continually protect that connection against changes in the physical environment, and movement ensures that those changes are unavoidable.

The timing of the decision matters as much as the decision itself. Transfer too late and the original radio connection may deteriorate before the handover can be completed. Transfer unnecessarily and the network consumes resources coordinating movement that provided little benefit. A subscriber traveling near the boundary between cells could encounter rapidly changing radio conditions, making the choice still more delicate.

The clean boundaries shown on cellular diagrams disappear entirely at this point. A vehicle does not cross an invisible line on the road where one cell ends and another begins. Coverage overlaps, signal strength varies, and the best serving cell emerges from the radio conditions rather than from a geometric boundary.

Handover is therefore where the abstract cellular map encounters the physical world. It also demonstrates why mobility cannot be solved by the handset alone. The mobile participates in the radio connection, but the network owns the larger problem. Base stations, radio resources, and switching infrastructure must cooperate to preserve a call as its point of radio access changes.

The subscriber experiences none of this as a new telephone call. There is no need to redial the number each time the car enters another cell, nor should the person at the other end need to know that anything changed. When the system works well, movement through the network becomes invisible. That invisibility is one of cellular telephony's most important achievements. Yet moving between cells inside one network is only one kind of mobility. A subscriber can travel much farther than the boundaries of a single cell, eventually leaving the area operated by the network with which they normally receive service.

Maintaining a call across cells requires cooperation inside a network—providing service beyond that network requires cooperation between networks.

Roaming

> **What happens when a mobile subscriber travels beyond the network in which their service normally belongs?**

Moving between cells is already a coordination problem, but at least those cells can belong to the same network. They can operate under the same administration, share infrastructure, and follow the same operational rules. Crossing into another network changes the nature of the problem because the subscriber is now asking one operator to provide service on behalf of another.

For mobile telephony to become genuinely useful across large geographic areas, the network could not treat every administrative boundary as the end of mobility. This was particularly important in the Nordic countries. People routinely traveled between neighboring countries for business and leisure, and a mobile telephone that stopped working at the border would preserve one of the limitations that cellular technology was supposed to overcome. The ambition behind NMT therefore extended beyond providing cellular coverage within individual national networks. Compatible networks could cooperate so that subscribers could continue receiving service while traveling within the participating region.

That capability became known as **roaming**.

Roaming means that a subscriber belonging to one network can receive service through another compatible network while away from their home network. The visited network provides the local radio access and cellular infrastructure, while the subscriber retains a relationship with the operator to which they normally belong.

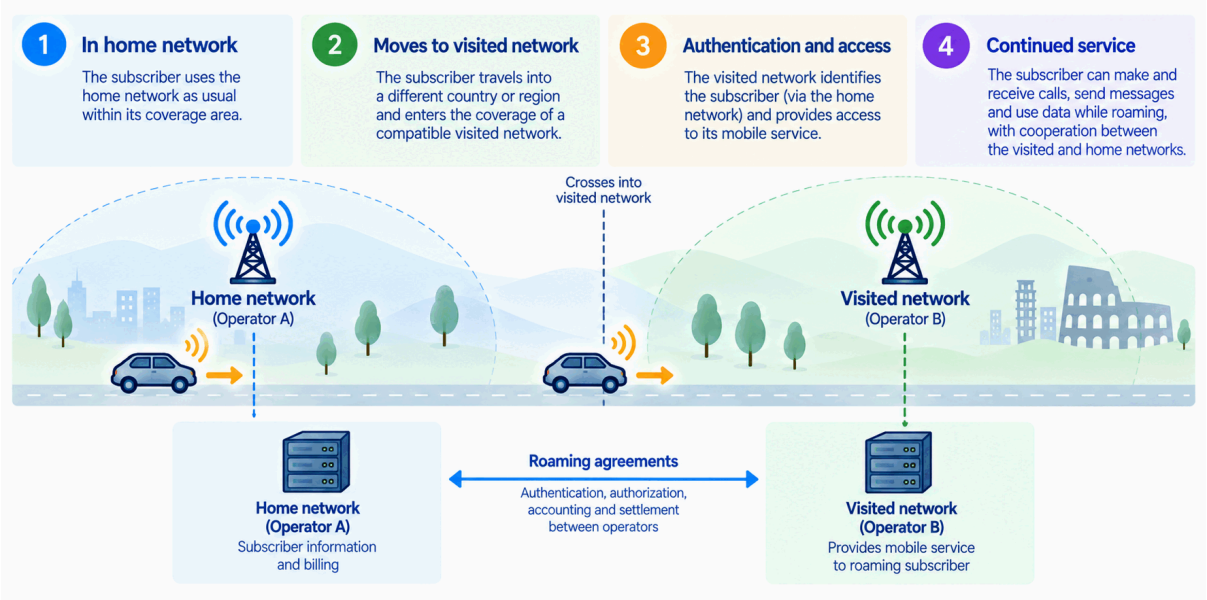


Figure 2.14. *A roaming subscriber leaves the coverage of the home network and receives mobile service through a compatible visited network, requiring cooperation between the two operators.*

At first glance, this might appear to be mostly a radio problem.

> **If the telephone supports the frequencies used by the visited network and understands the same cellular system, why should it matter which operator owns the base station?**

The difficulty is that receiving a radio signal is not the same as being entitled to use the network. The visited operator needs some way to recognize that the subscriber belongs elsewhere and determine whether service should be provided. The networks need agreed procedures for handling that subscriber, and the operators need a commercial relationship governing what happens when one company's infrastructure is used by another company's customer.

Roaming is therefore both a technical capability and an agreement between organizations.

That distinction is fundamental. Two networks can use compatible technology without automatically allowing their subscribers to roam between them. Conversely, a roaming arrangement cannot make fundamentally incompatible radio systems work together. The device, networks, operational procedures, and commercial relationships all have to align.

NMT provided an important early demonstration of what such cooperation could achieve. Its Nordic foundations meant that mobility could be considered across national networks rather than only within them. A subscriber could travel into another participating country and still use mobile telephone service,

turning the device into something associated increasingly with the subscriber rather than with one geographic network.

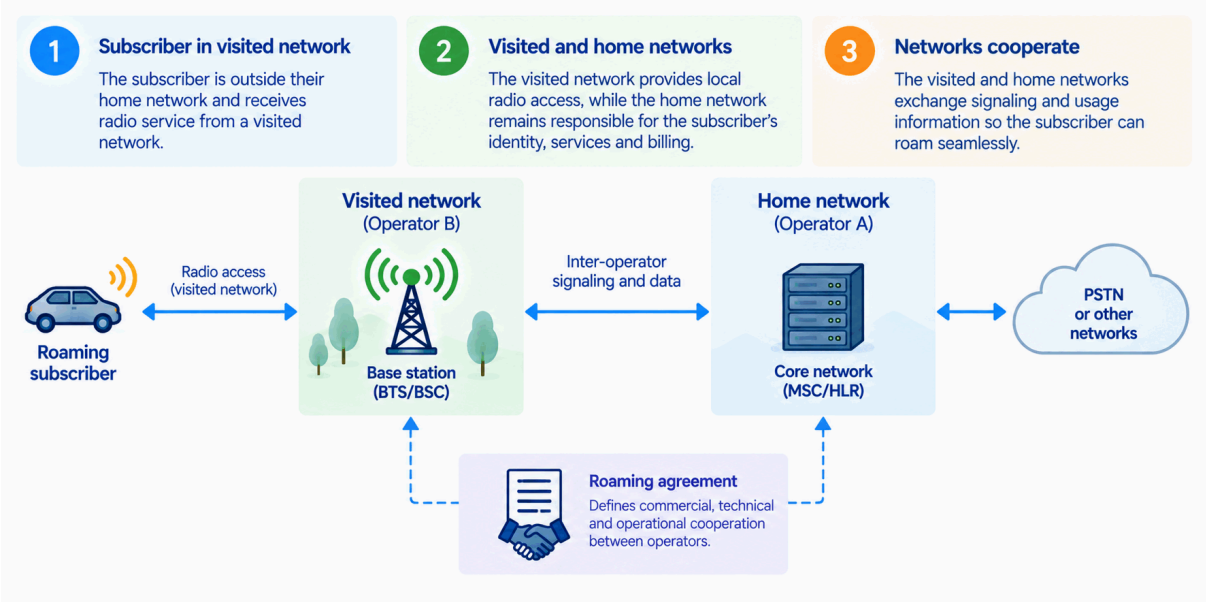


Figure 2.15. *Roaming separates the subscriber's home relationship from the network currently providing radio service. The visited network supplies local access while cooperating with the subscriber's home network.*

This introduced questions that would become progressively more important as mobile networks expanded.

> **How does the visited network recognize the subscriber?**
>
> **What information must be exchanged between operators?**
>
> **How should incoming calls reach someone who is no longer inside their home network?**
>
> **Who records the service being used, and how is the resulting cost settled between operators?**

The exact mechanisms would evolve substantially across later generations, but the underlying problem had already appeared: mobility could no longer be understood purely as movement between radio cells. A subscriber could move between **networks**.

ⓘ **Did You Know? — Roaming**
Roaming is often experienced as a feature of the phone, but it depends on cooperation beyond the device itself. Compatible radio technology is only part of the solution; operators must also agree on how visiting subscribers are recognized, served, and accounted for.

This makes roaming one of the clearest examples of why a mobile network is larger than its technology. Radio engineers can make two devices communicate. Network engineers can create mechanisms for mobility and call routing. But providing a service across organizational boundaries also requires operators to trust one another sufficiently to accept subscribers and account for their use of the network.

For the subscriber, the achievement was much simpler. The mobile telephone could continue to be useful even after the journey crossed an administrative boundary. The network was beginning to follow the person rather than forcing the person to remain within one network.

That was a remarkable expansion of what mobile telephony could mean. But every expansion exposed the limitations of the technology more clearly. More subscribers meant greater pressure on the available radio channels, while the analog radio connection itself provided protections that would soon prove inadequate for a service becoming increasingly valuable.

The first cellular networks had demonstrated that mobility could scale across cells and even across networks. Their success would make their remaining weaknesses much harder to ignore.

Capacity and Security Limitations

The first generation of cellular networks had solved a remarkable collection of problems. A subscriber could make and receive calls while moving, the network could transfer an active conversation between cells, and compatible operators could even cooperate to provide service beyond the subscriber's home network.

Success, however, created a new problem.

As more people wanted mobile telephones, more people began competing for the same finite radio resources. Cellular architecture had made spectrum reuse possible, but it had not made spectrum unlimited. Each cell could support only the traffic permitted by the channels available there, and every active conversation occupied radio resources for the duration of the call.

A network designed for a relatively small population could therefore encounter increasing pressure as mobile telephony became more popular.

The problem was especially visible wherever subscribers concentrated geographically. A cell covering a busy commercial district might experience far greater demand than one serving a sparsely populated area, even if both had similar radio resources available. The network could add infrastructure,

adjust cell sizes, and reuse frequencies more aggressively, but each improvement involved cost, planning, and another encounter with interference.

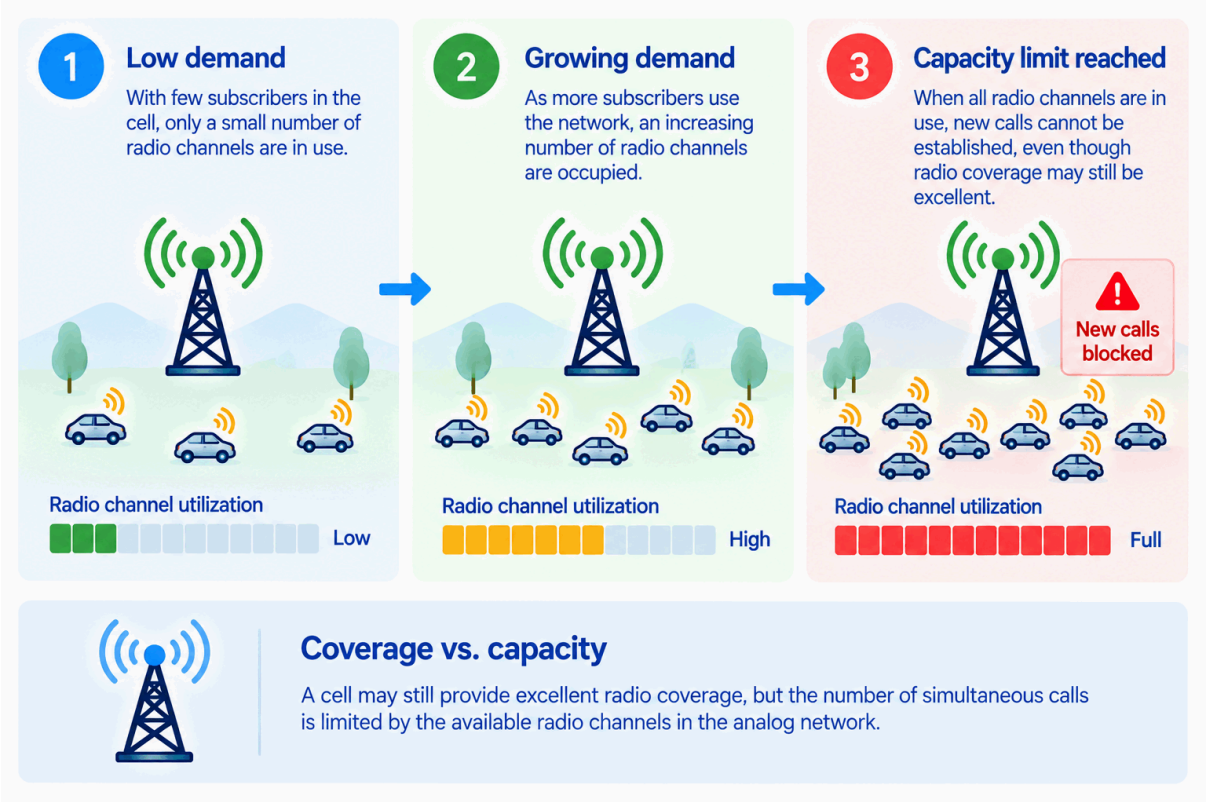


Figure 2.16. *As subscriber demand grows within a cell, an analog cellular network eventually reaches the limit imposed by its available radio channels, even though radio coverage may remain excellent.*

This exposed a limitation that had been present from the beginning: the cellular concept had improved capacity primarily by allowing frequencies to be reused across space, but it had not fundamentally changed the fact that an active analog voice call consumed a radio channel. Once the practical opportunities for adding cells and reusing frequencies had been exploited, further growth demanded more efficient ways of using the spectrum itself.

Capacity was not the only concern.

The radio link had another property that became increasingly uncomfortable as mobile telephony moved from an engineering novelty toward an important communications service: the signal was being transmitted through the air.

A wire entering a home provides a degree of physical isolation simply because someone normally needs access to that wire to observe the signal traveling through it. Radio behaves differently. A transmission intended for one mobile subscriber spreads through the surrounding environment and can potentially be received by other equipment within range.

The network may know which subscriber a channel has been assigned to, but the laws of radio propagation do not restrict the signal to that subscriber's receiver.

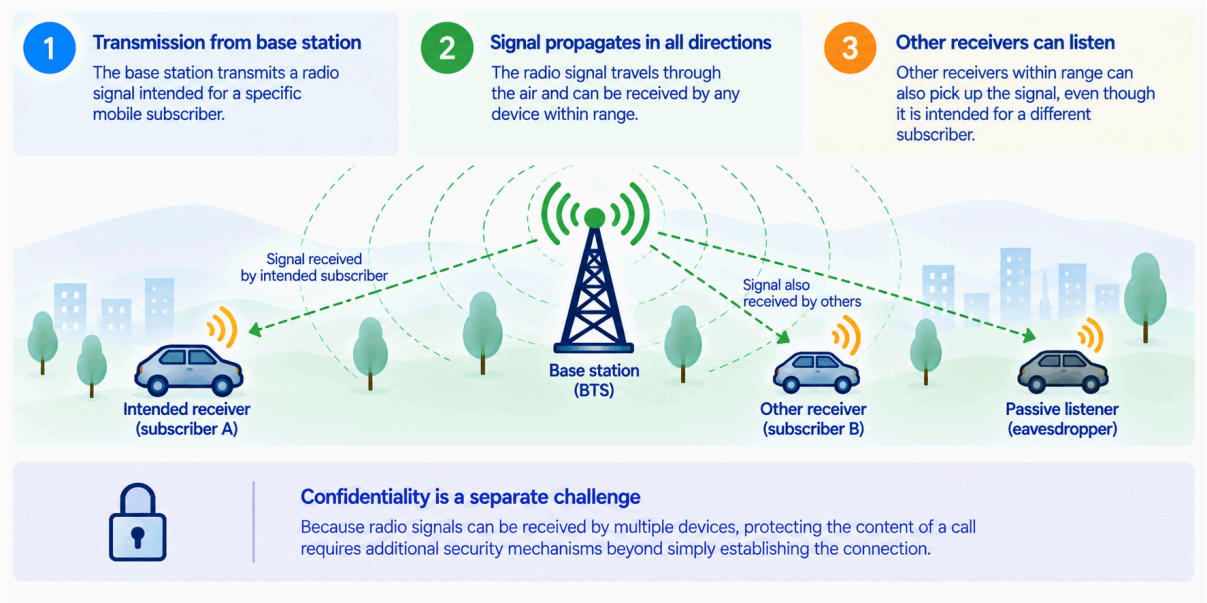


Figure 2.17. *A radio transmission intended for one mobile subscriber can also propagate toward other receivers within range, making confidentiality a separate problem from establishing the connection itself.*

For analog cellular voice, this created an obvious weakness. If someone possessed suitable receiving equipment and could tune to the relevant transmission, the fact that the network had assigned the channel to a particular call did not itself make the conversation confidential.

The distinction is important because **access control and confidentiality are different problems**. A network can decide which subscriber is permitted to use its service while still transmitting information in a form that another receiver may be able to observe. Protecting the network from unauthorized use and protecting the conversation from interception require mechanisms specifically designed for those purposes.

As cellular service became more commercially valuable, subscriber identity created another reason for concern. The network needed enough information to recognize customers and determine whether service should be provided. If those mechanisms could be observed, reproduced, or otherwise exploited, unauthorized users could potentially consume service while appearing to the network as legitimate subscribers.

These weaknesses reveal where successful analog cellular architecture was beginning to encounter the limits of its implementation.

ⓘ Did You Know? — Call Privacy

The fact that a cellular call had been correctly established did not mean that the radio transmission itself was private. In an analog system, controlling who was allowed to use the network and preventing someone else from listening to the radio signal were separate engineering challenges.

Neither problem diminished what 1G had achieved. The first commercial cellular systems had demonstrated that large-scale mobile telephony was possible. Their limitations became important precisely because people wanted more from them: more subscribers, more capacity, greater privacy, stronger protection against unauthorized use, and eventually services beyond an analog voice conversation.

The pressure was therefore no longer simply to build more cellular infrastructure.

Engineers needed to extract more capability from the radio spectrum and gain greater control over the information being carried across it. Voice could no longer remain merely a continuously varying signal transported over an analog radio channel if mobile networks were to satisfy those growing demands.

A different representation of information would open possibilities that analog cellular radio could not easily provide—the mobile network was ready to become digital.

Looking Ahead

The first commercial cellular networks proved that mobile telephony could become much more than a specialized radio service.

Cells provided coverage without requiring a single transmitter to dominate an entire region. Frequency reuse allowed scarce spectrum to support far more conversations than earlier mobile systems could accommodate. Base stations connected moving subscribers into switching infrastructure, while call-control mechanisms allocated channels and established connections through the public telephone network. Handover allowed those connections to survive movement, and roaming demonstrated that mobility could extend beyond the boundaries of a single network.

Together, these capabilities transformed an engineering concept into a practical service. More importantly, they changed what people could reasonably expect from a telephone. The connection no longer belonged to a building or a particular pair of wires. A subscriber could travel through a city, move between cells, and even enter another compatible network while remaining part of the telephone system.

Yet the success of cellular telephony placed increasing pressure on the very technology that had made it possible. Every additional subscriber represented another potential demand for radio resources. Every active analog conversation occupied part of the finite spectrum available within a cell. Operators could deploy more infrastructure and reuse frequencies more frequently, but there were practical limits to how far capacity could be increased through geography alone.

The radio connection itself presented another set of limitations. Analog voice could be transmitted effectively, but the network had comparatively little control over the information once it was being carried through the air. A receiver within range could potentially observe a transmission, while protecting subscriber identity and preventing unauthorized use demanded stronger mechanisms than the earliest cellular systems had been designed to provide.

The problem was no longer whether cellular telephony worked. It worked well enough that its limitations mattered. A growing mobile service needed to carry more conversations within the available spectrum, provide stronger mechanisms for protecting subscribers and network access, and support an ecosystem that was becoming increasingly international. The telephone experience might remain familiar, but the representation of information underneath it did not have to remain analog.

That possibility changes the engineering problem considerably. If speech can be represented as data rather than carried directly as a continuously varying analog signal, the network gains new ways to process it before transmission. Information can be encoded, protected, interleaved, encrypted, and reconstructed. Radio resources can be shared using techniques that are difficult to achieve when the network is primarily concerned with preserving an analog waveform.

The significance extends beyond voice.

Once a mobile network learns to transport digital information, the radio connection is no longer inherently tied to the sound of a telephone conversation. Voice becomes one form of information among others, opening the possibility that the same network might eventually carry messages, signaling, and services that have little resemblance to a traditional telephone call.

The cellular architecture did not need to be abandoned to make this possible. Its most important principles—cells, frequency reuse, mobility, switching, subscriber management, and roaming—had already demonstrated their value. The opportunity was to preserve those ideas while changing the way

information traveled across the radio network. This reflects the broader progression: mobile-network evolution is cumulative, with later systems inheriting architectural principles while changing the technologies used to implement them.

Analog cellular networks had made the telephone mobile. Digital technology would begin transforming what that mobile network could become.

PREVIEW EDITION

to purchase a copy of the book
visit the following website

www.ardiri.com

the book is available in both
digital and physical form.

Every connection hides a network.

A phone finds a signal.
A message crosses the world.
The Internet appears almost instantly.

Yet behind those simple moments lies one of the most complex systems ever engineered—a global network of radios, protocols, identities, switching systems, packet cores and carefully coordinated standards that has evolved over generations.

Inside the Mobile Network peels back those layers one by one, revealing how cellular networks really work. Beginning with the first analog systems and progressing through GSM, packet data, 3G, LTE and 5G, it follows the evolution of the technologies and architectures that transformed the mobile phone from a voice terminal into a gateway to a globally connected world.

Along the way, it explores not only **what** changed, but **why**—how engineers solved the problems of spectrum, mobility, authentication, roaming, packet data and interoperability, and how thousands of competing companies ultimately learned to build systems that could work together.

Whether you're a student, developer, engineer or simply curious about what really happens when your phone connects to a network, this book will change the way you think about those few bars at the top of your screen.

Because once you understand the network, you'll never look at a mobile phone in quite the same way again.

ABOUT THE AUTHOR



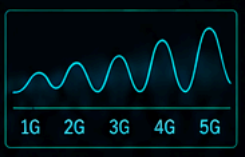
Aaron Ardiri is passionate about making complex technology accessible, he writes to help reafersunderstand not only how the mobile network works, but why they were designed the way they are.

1 1 1 0
0 1 0 1
0 1 1 1
1 1 0 0
0 1 0 0
0 1 1 0
1 1 1 0
0 1 0 0
1 1 1 0
1 0 1 1
1 1 0 1
0 1 0 1
0 1 0 1
1 1 1 0

"It's not just about mobile phones.
It's about the networks, technologies, standards and people that connect the world."

RADIO ACCESS NETWORKS

Cells, spectrum and wireless technologies from 1G to 5G.



CORE NETWORKS

Mobility, session management and services that keep the world connected.

AMF	SMF
UPF	PCF
AUSF	UDM
NRF	NEF

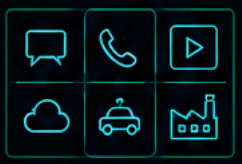
DEVICES & SIM

From the first mobile handsets to modern smartphones, and the identity that connects them.



SERVICES

Voice, messaging, mobile broadband, IoT, edge computing and beyond.



STANDARDS

Inside 3GPP — how thousands of experts create global standards that make mobile networks work.



SECURITY

Authentication, encryption and trust in a connected world.



WHAT'S NEXT?

5G Advanced, non-terrestrial networks, private networks, AI and the road to 6G.



ISBN 978-91-535-1698-9

1 0 1
0 1 0
1 0 1

FROM SPECTRUM TO SERVICES. FROM STANDARDS TO DEPLOYMENT. FROM CONNECTIONS TO OPPORTUNITIES.