# 2

# Mobile Network History

Brian Brown, Rob Gonzalez, and Brian Stanford
*Cisco Systems, Herndon, VA, USA*

## 2.1 Overview

Few things have changed the way we work, live, and play more than the evolution of mobile networking. In 1990, landlines were the big moneymaker for the service providers (SPs). Little did they know that a massive disruption in the marketplace was going to develop and in the middle would be mobile networking. Since the acceleration of this change through the decades, kids and teenagers today have never seen a pay phone, regularly carry a small computer (smartphone) in their pocket, and are connected close to 24 h a day. As technology continues to evolve, new areas of development are coming to the forefront of the discussion. Virtualization, orchestration, and scalability are now big concerns as more and more applications, data sources, and users around the world discover the possibilities. Network connectivity is critical to provide mobile users the services and experience they are looking for today (Fig. 2.1).

As the demand for mobile services has changed from the beginning of the commercially available GSM in the early 1990s to the evolution to a packet-based architecture in GPRS, to a more robust service in UMTS, and to the more mature designs in LTE and beyond, the supporting core architecture has also changed. In this chapter, we will discuss the history and evolution of the mobile network. The initial demands and drivers for a mobile network have also changed over the years. While voice was the primary service during the initial architectures, data and video have overwhelmingly dwarfed voice. What technology changes pushed the network to places unheard of in the early parts of the mobility movement? Why did the industry start where it did, and why did it end up where it has? How have the network owners changed over the years, and are the same services driving the push for progress? How have user expectations changed along with these technology changes?

*Software Defined Mobile Networks (SDMN): Beyond LTE Network Architecture*, First Edition.
Edited by Madhusanka Liyanage, Andrei Gurtov, and Mika Ylianttila.

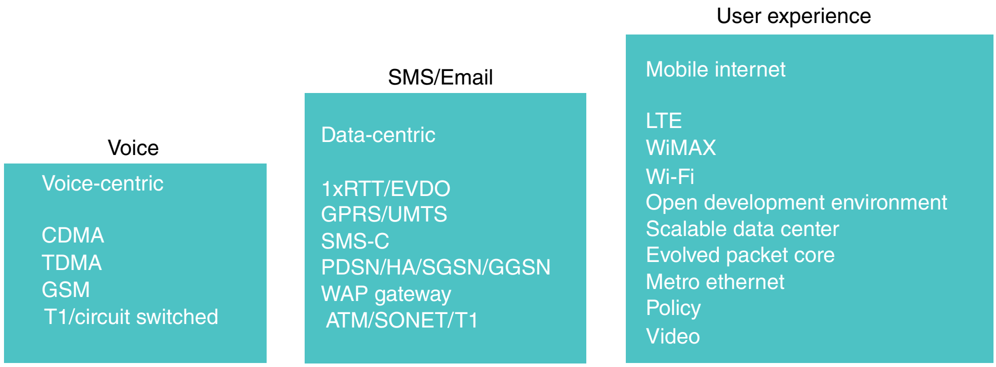


**Figure 2.1** Evolution of the mobile industry.

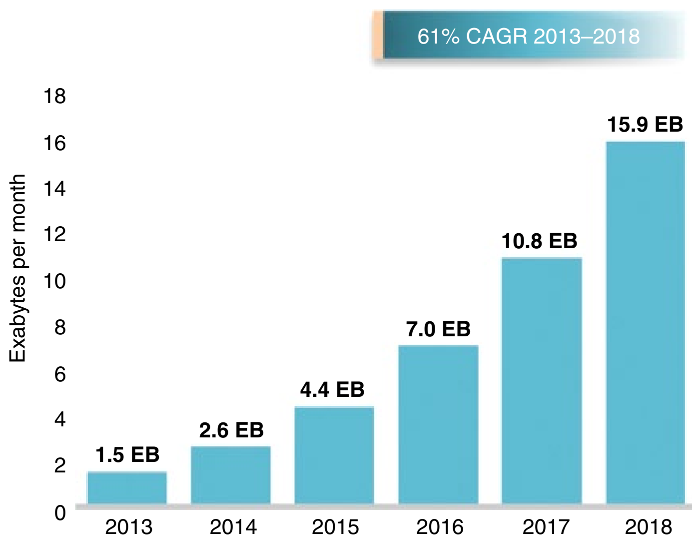


**Figure 2.2** Global mobile traffic per month. Source: Cisco Systems (2014) Cisco Visual Networking Index (VNI).

## 2.2 The Evolution of the Mobile Network

At the end of 2013, global mobile data traffic reached 1.5 exabytes per month [1] and is expected to reach 15.9 exabytes per month by 2018 (Fig. 2.2).

The network has come a long way from the days of the 56K leased lines and T1s along with the large, clunky mobile phone in the early 1990s. Today, as the mobile endpoints become more and more numerous, the amount of data required to serve these endpoints becomes an issue along with managing the network access of these mobile devices. In 2013, over half a billion (526 million) mobile devices and connections were added. The Internet of Things is contributing to the increase in mobile traffic led by the transition to smarter mobile devices, emergence of wearable devices, and the increase in machine-to-machine (M2M) connections.

Today, the new data center architecture touts virtualization, orchestration, and scalability. Are these new concepts for the mobile network and where did it all start?

### 2.2.1 Sharing Resources

Back in the 1980s, because voice was the main driver for all SP business, the fledgling data network, including the mobile data network, was modeled after the existing voice network. Voice was carried over dedicated time-division multiplexed (TDM) lines designed to tightly pack in the newly digitized voice traffic into 64-kilobit channels. If a user in Atlanta, Georgia, needed to communicate with a user in San Francisco, California, a dedicated voice channel was established in order to support the 64-kilobit voice call. This dedicated channel could be used in any way for the duration of the call but may be wasted if no one was speaking. The bandwidth was reserved for this connection until the end users disconnected. Users had become accustomed to high-quality voice services due to the high availability (HA) of the legacy voice networks. Today, users are more tolerant of voice anomalies like choppy voice and dropped calls as they have migrated to more data usage.

Today, the ideas of sharing compute and storage resources are not new concepts. They are an old concept applied to a more current technical architecture. The mobile SPs in the early 1990s used the existing technology of the day to share what resources they had between users in order to efficiently deliver voice services. In the United States, a provider could multiplex 28 T1s into a single T3 and further multiplex many T3s into larger optical circuits. As newer multiplexing technologies were introduced, the result of sharing those resources was that the cost of a single T1 continued to drop. Asynchronous Transfer Mode (ATM) was also used to share high bandwidth connections among many customers. These shared large-capacity trunks were used to pass thousands of customer circuits between two endpoints on a network. ATM provided the higher bandwidth and quality of service (QOS) needed for evolving networks that were transporting voice, video, and data. Additionally, once the infrastructure was installed, new customers were quickly and easily added relative to before.

As the network transitioned to more of a packet-based architecture and voice and data were converged onto the same network, sharing bandwidth now improved even more, enabling the ability to share more resources. Circuits were no longer used per user but could now carry multiple users and applications, further moving the architecture to a shared space. Where before, in the circuit-switched world, a dedicated circuit had to be established and maintained during a call, the new packet-based network could share bandwidth on a per-packet level. As we migrated to the packet-based network, other technology contributed to the efficiency of the overall architecture. For example, silence suppression on voice trunks was one feature that would not transmit packets if an end user were not speaking, thereby increasing the sharable bandwidth in the network.

The network infrastructure was designed with HA in mind, so SPs had to reserve bandwidth to account for network node or link failures. The addition of redundant paths and redundant nodes along with many protocols has increased the complexity of the network infrastructure. SPs operating a highly complex network have higher operational cost, and they tend to have less network availability. The sharing of network resources that reduced the cost of providing the converged infrastructure for the SPs has now put a higher priority on HA and QOS due to the impact of network outages. As user demands grow, SPs must find a cost-effective way to add more users to the existing infrastructure along with the ability to deploy new services. Along with these technology changes and user demands, SPs are finding it harder to maintain the same level of services while expanding their networks.

In addition to sharing transport resources, the mobile packet core was also in transition. Where before, dedicated resources had to be installed and maintained for every specific function of the mobile packet core, now these functions can be shared on the same computing platform.

### *2.2.2 Orchestration*

Wikipedia defines computing orchestration as “the automated arrangement, coordination, and management of complex computer systems, middleware, and services.” This term is now widely used when talking about the cloud network. In layman’s terms, orchestration is the higher-level coordination of hardware and software components, stitched together to support a particular service or application, and the active management and monitoring of that network. Again, this is not a new concept; however, technology today enables companies to automate the process in ways unheard of several years ago. How far back can we go to witness the beginning of the orchestration concept in action in the fixed or mobile communication system?

In the early days of telephony, a phone call was initiated by picking up the receiver and telling the operator the number that you wanted to connect. The operator then physically connected a wire to the next hop of the circuit. Since most calls were local, the operator connected the two callers together. When the call was completed, the operator had to disconnect the circuit. Could you imagine having to tell an operator each time you wanted to go to a new Web site? Fortunately, we have automated many of the processes that used to be done manually. This is the journey of orchestration.

In the early days of commercial mobile services, mobile SPs held a monopoly on the network, the devices connected to the network, and the applications that ran over the network. Development of new applications took months, if not years, to create. When a new application was ready to be introduced into the network, the orchestration of that implementation was affected by several factors.

Computing technology was still being developed, and all documentation was still largely being delivered using paper copies of service orders and directives. This communication started at the headquarters location and was delivered to any and all field locations with tasks to complete. As each field location completed its tasks, it communicated this back to the headquarters. After the field locations were completed, the service was then tested. Once testing was completed, the application could be put into service. This process was thorough, but not fast. Communication was slow, network visibility was limited, and the motivation to quickly implement the service was lacking. Competition was not yet a driver for improvements in this process.

Today, with the proliferation of cheap memory, large database capabilities, fast network communications, virtualized computing, and standard connections, the scenario above can be automated and completed in a fraction of the time. The concept is still the same. A central figure dictates what needs to be done at each location in the network. Once the completion has been acknowledged, the service or application can now be put into service.

### 2.2.3 Scalability

As mobility services grew in popularity, the SPs needed to address scalability issues. These issues included expanding points of presence (POPs) to facilitate additional equipment to terminate more users and additional circuits for core capacity along with determining where to connect these new core links. The early models used to design voice networks continued to be used as mobile voice services grew. Traditional point-to-point circuits connecting early services (2G/3G) to simple hub-and-spoke architectures were the norm. As data services grew, the design model had to evolve to provide cost-effective bandwidth in the network—evolution to a mixed architecture of hub-and-spoke architecture for S1 interfaces and meshed architecture for X2 interfaces. The outcome of the changes with network design provided ample bandwidth in the core, so QOS was not addressed in this area. The access into the network from the edge was all that was needed to address QOS. As data usage grew, providing the customer with a good experience had to be addressed. Due to the high cost of core bandwidth, it quickly became cost prohibitive to keep adding enough capacity to handle peak data load in the core. Network modeling evolved to include a more complex QOS strategy, which addressed the increased cost of adding additional bandwidth. As SPs have upgraded their core network technology from TDM, ATM, POS, and Ethernet for transport, additional technology has evolved to provide path selection along with guarantees to the user traffic carried on SP networks. SPs want to ensure that signaling traffic takes the shortest path in order to preserve both voice calls and data sessions as users are moving, which require the voice call or data session to switching from one base station to another. Scalability design usually involves a trade-off of maximum size of the network elements (the number of links, nodes, and protocols) and minimum availability. SPs need to ensure that the network will remain stable in order to provide the best user experience but also need to expand the network. This will continue to be one of the main challenges with the evolving mobility networks.

## 2.3 Limitations and Challenges in Current Mobile Networks

With the introduction of the smartphone and the consumer's ability to change the operating system and choose their own applications, customers achieved true control of their personal devices for the first time. Prior to this, the control consumers wielded was limited to the decision of which phone to buy only. After making that fateful decision, they were constrained to use the application offered on that phone. If customers wanted a navigation application, they were forced to pay whatever fee the phone provider wanted to charge for that application.

Happily, this is no longer the case. Now, the consumer can pay a small one-time fee for an application or use one of thousands of free applications. Choices in devices also increased with not only phones but tablets, computers, cellular modems, and personal hotspots to share cellular connections. The smartphone is no longer a mere toy for the personal sphere; it is an essential business tool, and there are many choices depending on the needs and personal preferences of the consumer. But with this growth in choice comes a range of additional pressures on SPs who must support their customer's evolving tastes and behaviors. Smartphones require more data and are expected to be 96% of the global mobile traffic by 2018 [1] (Fig. 2.3).

Customers now use the network in different ways and not just for phone calls and SMS messaging. We are sending email, surfing the Web, accessing private corporate networks, editing

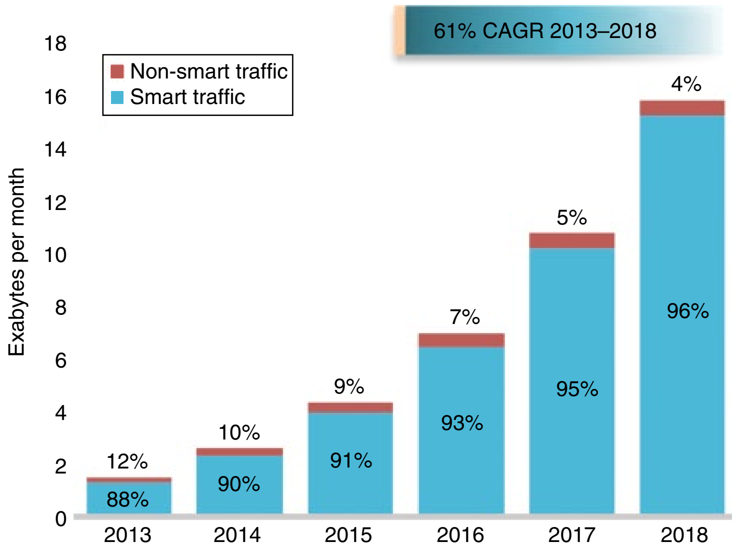


**Figure 2.3** Effect of smart mobile devices and connections growth on traffic. Smart traffic is generated by devices with advanced computing capabilities and a 3G connection (or higher). Source: Cisco Systems (2014) Cisco Visual Networking Index (VNI).

documents, listening to music, watching video, and even watching live TV. Some of us hold our digital lives in these devices with everything from the personal and work spheres coexisting: family photos and video, the family calendar, corporate email, and confidential documents of all stripes. The user is also more mobile than ever, and we expect a similar experience on our smartphones to that of sitting at our desk in the office. With mobile devices, the business user can have even more now with the use of virtual desktop infrastructure (VDI) that allows them to access a virtual intestine of their desktop with secure access to the corporate network and applications. Compounding pressures exerted by user expectations for performance and flexibility are yet other phenomena in the new landscape of the technology world.

Over the top (OTT) refers to video, music, and other services provided over the Internet rather than being provided by the mobile SP. OTT providers such as Amazon Prime, Netflix, Hulu, iTunes, and numerous other content applications are using large amounts of data and are not under the provider's control. OTT traffic accounts for the majority of mobile Internet traffic today as it is continuing to grow. Some of these services compete directly with the mobile provider and affect the potential revenue as well as drive the continual need for more bandwidth.

The old world of primarily voice and SMS applications and for which much of the current network apparatus was built requires far less bandwidth compared with the new video-intensive applications that are far more "bursty" in nature. This type of traffic is bandwidth intensive depending on what the user is doing and is latency sensitive. If the latency is too large or variable, the user experience will be reduced and can possibly drive the user to find other means of working including switching providers. SPs must find a way to adapt to the new wave of users but can't neglect older services either.

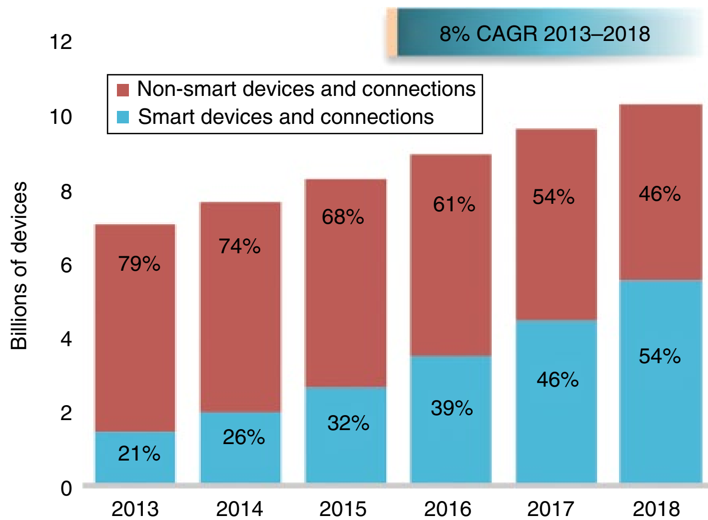


**Figure 2.4** Global growth of smart mobile devices and connections. Smart devices are those having advanced computing capabilities with a minimum of 3G connectivity. Source: Cisco Systems (2014) Cisco Visual Networking Index (VNI).

With the rapid evolution of the mobile phone systems, mobile providers need to maintain multiple networks to support all of their 2G, 3G, and LTE services. Though a large percentage of users are migrating to new LTE-enabled smartphones [1], there still exists a significant block of older technology consumers that demands the care and feeding of multiple networks (Fig. 2.4).

While some reuse of components can be accomplished, the demands of the "new network" require the insertion or overlay of a number of new devices and technologies. This leads to a proliferation of support systems (e.g., OSS/BSS/NMS) required to keep these systems functional.

Providers also must balance the need to have capacity to cover current term needs with the costs of maintaining unused bandwidth. The data usage increase has led to carriers needing to increase backhaul capacity with various high-speed connections including metro Ethernet and dark fiber. The time to add capacity can be days to months and varies by type of connections as well as whether they own or lease the circuits. Adding backhaul circuits can also require multiple truck rolls to install, provision, and test new circuits. While mobile carriers are needing to upgrade their networks to handle increased traffic, they are also seeing a reduction in the price of voice minutes, declining text messaging, and loss of revenue to OTT services. At the same time customers are demanding more data bandwidth and lower costs. The challenges continue to increase for the SP.

Looming on the horizon are the anticipated growth and corresponding demands on the underlying networks and systems expected from the advent of the Internet of Things. An anticipated explosion in connectivity driven by everything from sensors on cars and other inanimate objects to health information from body sensors could overwhelm the existing infrastructure. Mobile-to-mobile connections will grow from 341 million in 2013 to over 2 billion by 2018 [1] (Fig. 2.5).

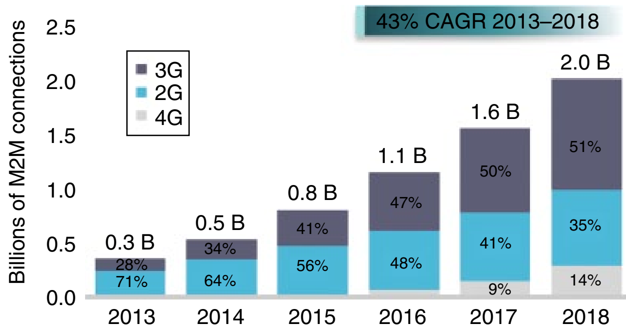


**Figure 2.5** Global machine-to-machine growth and migration from 2G to 3G and 4G. Source: Cisco Systems (2014) Cisco Visual Networking Index (VNI).

Admittedly, many of these devices require low bandwidth, but who can foresee the full growth and need as more and more devices not yet being considered are added to the constellation of devices being added each day? History tells us that no one truly can, but we only know that there will be surprises.

## 2.4 Requirement in Future Mobile Networks

Understanding the challenges of today and looking at the trends of the future, it is clear that the way mobile networks are managed and built needs to change. There is the need for better orchestration as well as in-depth analytics to ensure the successful operations of mobile networks in the future. Monthly global mobile data traffic will surpass 15 exabytes per month by 2018 from close to 10 billion mobile devices and connections [1].

Mobile providers will have to adapt to the growing number of devices and traffic demands. To do this, they need to be able to provision, manage, and optimize the mobile network end to end. Using common tools and standardized application programming interfaces (API) to communicate with devices will remove a great deal of network complexity. This allows providers to quickly and efficiently apply new or updated service policies to their evolving network's traffic engineering needs at any point in the network.

An increasing number of mobile users will connect from a fixed location. This fixed nature places an increasing demand in areas of concentrated populations. Users will demand a seamless transition between fixed and mobile connections. With increasing mobile data rates available to users, mobile providers will become a wireline broadband competitor. LTE networks growth may grow faster as mobile users continue to demand similar service from both fixed and mobile networks. To assist with more fixed users as well as to deal with areas of dense population of mobile users, mobile providers are looking at small cells. Small cells have limited range but use low power and can be deployed in congested areas such as stadiums, city downtown areas, auditoriums, and neighborhoods to off-load mobile data traffic as providers deal with the large growth in traffic. There are forecasts for millions of small cells in the near future and adding to the complexity of the network that needs to be operated. Mobile off-load traffic is seen to increase from 1.2 exabytes/month in 2013 to 17.3 exabytes/month by 2018 [1] (Fig. 2.6).

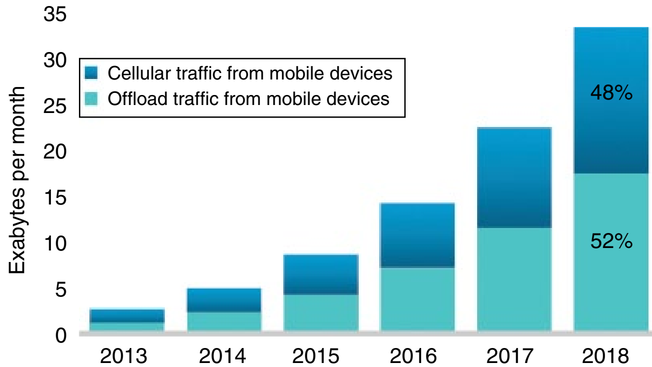


**Figure 2.6** Fifty-two percent of total mobile data traffic will be off-loaded by 2018. Source: Cisco Systems (2014) Cisco Visual Networking Index (VNI).

The need to extract data from the network is becoming a must for mobile providers—from analyzing and even predicting usage patterns to location-based services. With location-based services, the mobile provider can work with stores to provide customers with coupons as they enter the store or provide offers for new services including Wi-Fi access for repeat customers. For this to be effective, mobile providers must be able to quickly extract the network inelegance, process the data to determine if the store or merchant has any offers, and then present the end user with the offer. This will require not only the ability to collect the data but also the computing resources to identify the needed information and act of that information.

Networks are already complex to manage, but with the addition of possibility of millions of small cells in the near future, this becomes even more challenging and impossible to optimize manually. Extracting information from the network will also allow mobile providers to apply business intelligence to optimize the network.

Without a standard API to collect information from the network and a robust orchestration system to automate the changes, the network will not be able to respond to the traffic demands. The standards will also reduce the case of vendor lock-in due to proprietary features or proprietary management systems. Operators need to be able to have a high-level control point where they can quickly apply changes to the network as a whole and manage dense deployments.

Without this simplicity, operators will continue to struggle with very large complex networks that make deployment more time consuming at a time when networks need to be more flexible to meet the ever-changing demands of end users.

## Reference

[1] Cisco Systems (2014) Cisco Visual Networking Index (VNI). http://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html (accessed February 17, 2015).