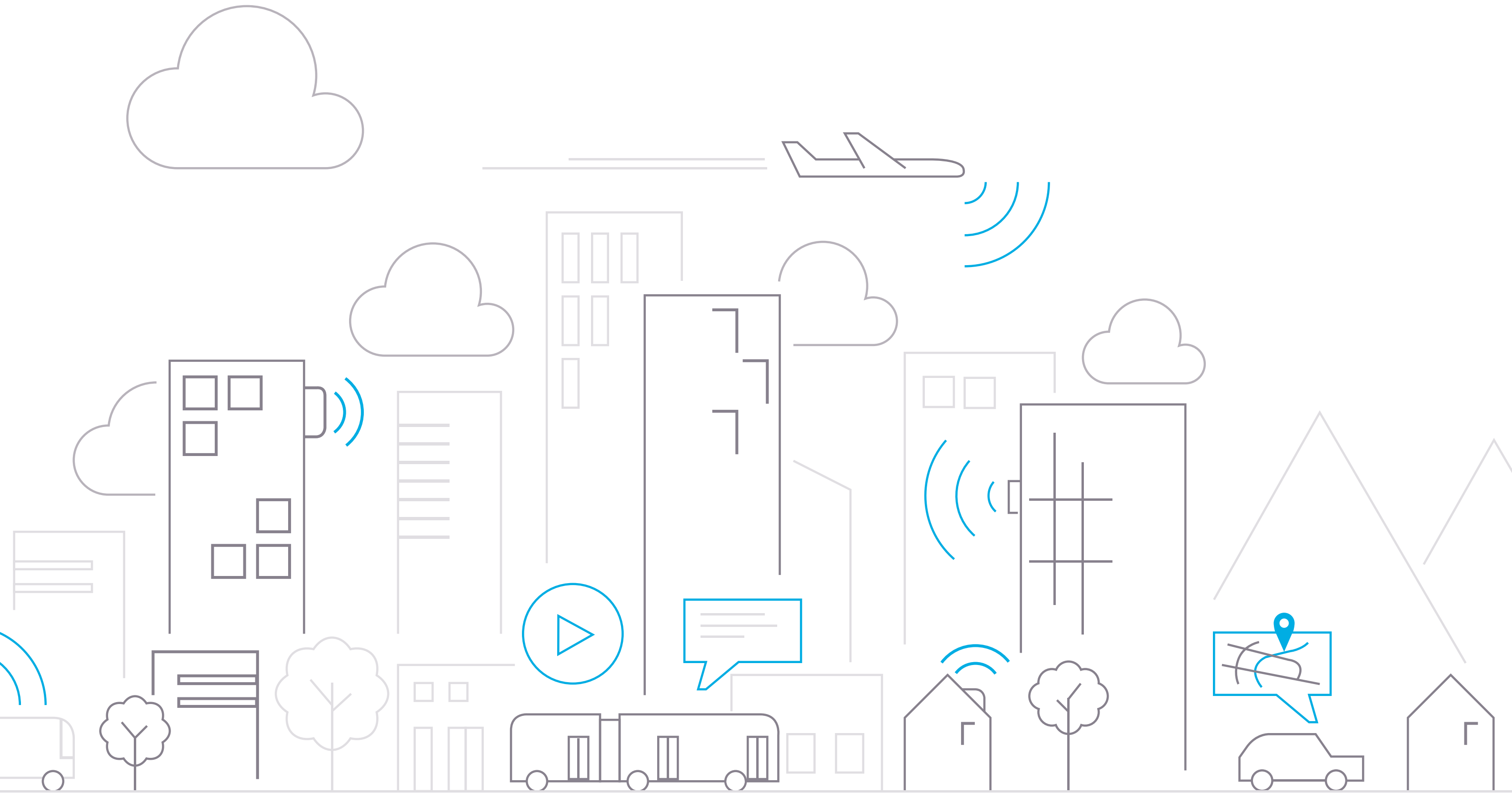
Bell Labs Consulting



Who will satisfy the desire to consume?

How will wireless data demand grow between now and 2020 and how should mobile operators respond?

NOKIA

About this report

This Bell Labs Consulting report offers mobile operators and other industry stakeholders a unique perspective on demand for mobile data through 2020 with a new approach to modeling the demand growth. By concentrating on people's desire for digital content and services rather than mobile traffic trends, this report reveals just how much potential demand is available for mobile operators to address.

How to address that demand forms the second part of the report. Recognizing that mobile operators are constrained by business concerns and technology limits, Bell Labs Consulting models show the kinds of transformations that will be required to unlock the future opportunity — helping operators weigh their choices about when to evolve, by how much and at what cost with a high level of confidence.

Contents

It's all about the network

The next evolution of our digital global civilization will be driven by life and business automation in a world in which billions of interconnected things — smart devices, objects and sensors — exchange information that is used to create new knowledge which, in turn, will be used to automate mundane tasks. Automation will make the systems and processes that underpin business and consumer lives more efficient and more intelligent. It will also change human behaviors, by further increasing our dependence on the underlying network to provide connectivity that will allow us to communicate, collaborate, consume content, control systems, processes, and devices, all in a completely contextual way.

Seamlessness and contextual communications, collaboration and control of and with everything and everyone will be of critical importance in the future. Communications will move fully into the "post-voice" era in which interactions shift effortlessly between people, apps, systems and Internet of Things (IoT) devices using chat, video, voice, gestures and more.

And future enterprise environments will provide the same flexible, agile, seamless and mobile experience that digitally sophisticated consumers are used to in their personal lives, allowing business to be conducted without interruption, anywhere, anytime. Part of what will enable this will be the adoption of software-defined networking (SDN) in the data center, the LAN and the WAN.

All of these trends will produce a dramatic shift in demand — challenging mobile operators to provide networks and platforms that achieve the highest performance at the lowest cost per bit while supporting extensive personalization.

The network will require massive elasticity and scalability to match changing demand of new services and applications. Moreover, new highly interactive and time-critical control and human perception-dependent applications (for example, virtual reality/augmented reality) will require ultra-low latency. The best, and in many cases, the only way to achieve this scalability and low latency is to move the processing and rendering to the very edge of the network, as close to the end users as possible and to dynamically optimize transport paths for data delivery at the lowest cost per bit.

This inaugural Bell Labs Consulting report looks at what all of the above means for today's mobile operators. It does so in a unique way, reflecting the long tradition at Nokia Bell Labs of turning conventional questions around to uncover new insights and answers.

Rather than extrapolate future mobile traffic demands based on current baselines and growth rates, this report presents a new demand model built from the ground up, based on our own research and external data, offering important insights into the future. It addresses three key questions:

1. How big is the potential demand for new services today and by how much will it grow by the year 2020 — globally and regionally?
2. How much of that demand can be met by unlicensed spectrum solutions (for example, Wi-Fi) and how can mobile operators profitably deliver the remaining (high mobility, long range, high performance) services?
3. How does the network have to change — how can new technologies and architectures help address the critical challenges by dramatically lowering costs and/or improving performance, adaptability and programmability and enabling new services and business models?

Taken together, these questions — and their answers — start a critical conversation about the future of networks and how to build them.

Marcus Weldon
President of Nokia Bell Labs

How would users behave if costs were no longer a constraint, if Wi-Fi was ubiquitous and cellular (licensed) networks were massively scalable with optimized economics?

What if consumers and businesses had the same anywhere, anytime access, to convenient affordable technologies that home users do?

How much pent-up and future demand will there be for network operators to address and how can they address it with sustainable economics?

Understanding the demand

Growth in mobile data traffic continues unabated, as do analysts' forecasts of this growth. Most forecasts focus on traffic measurements and trend analysis based on extrapolations of current traffic trends, which are of limited value in preparing for the future that is markedly different from the past and present. Traffic observed today is a function not just of the demand for digital services, but also their affordability. Affordability, in turn, depends on the necessary profit margins, cost structures and business models of operators. As new technologies continue to significantly improve cost structures and enable new services and business models, we believe that it is possible to derive new insights into the future by first modeling the true potential demand as if the affordability constraint were removed.

Based on this approach, Bell Labs Consulting estimates a global average increase in demand of 30 times to 45 times from 2014 levels — with some markets experiencing as much as a 98 times jump over the same period.

Wi-Fi is cost effective due to its use of unlicensed spectrum and simple access and control mechanisms. Wi-Fi service is readily available in developed countries today and is expected to grow significantly in developing parts of the world. For example, in North America (NA) alone there are over 100 million access points deployed. By 2020 we predict that Wi-Fi will have the potential to satisfy two-thirds of all growth in demand globally — more in some of the developed markets and less in other regions.

The remainder of the massive demand increase will need to be addressed by cellular (mobile) technologies. The key question is how can operators of such networks address this demand profitably? What new networking technologies and architectures will be required? Answering these questions will allow us to find the equilibrium between demand and supply and to understand whether the future is attainable or not.

Fine-tuning the look forward

While making projections based on current traffic trends is of limited value, trying to predict exactly what applications will emerge as the most popular ones to drive traffic in the future is virtually impossible. For example, in 2009 there was no concept that SMS would be replaced by new messaging platforms generating 4 times the number of messages, or that streaming video services would form 75 percent of internet traffic. While emergence of specific applications cannot be predicted, potential demand for applications such as messaging and video streaming — if the affordability constraint is removed — is more predictable. Hence, instead of trying to anticipate what new applications might develop in the next few years, Bell Labs Consulting has changed the question. If users are behaving a certain way today, how are they likely to behave in the next 5 to 10 years? We model the likely evolution of our usage of applications using five all-encompassing "canonical" categories: streaming, computing, gaming, communications and storage applications. We posit that all services can be mapped into these applications; for example, a virtual reality application is a combination of a computing and streaming application, and a video calling/chat application is a combination of a communications and a two-way video streaming application.

Removing the affordability constraint on usage allows us to model a world where universally affordable, high quality and seamless data services are readily available wherever and whenever needed. This unconstrained potential demand is modeled by accounting for consumption behavior of various user segments for the five application categories. This is cross-correlated for different locations, preferred devices and different times during the day. The demand potentially addressed by the projected availability of Wi-Fi (or Wi-Fi-like technologies such as MuLTEfire that use unlicensed spectrum) is then subtracted to arrive at the demand forecast that is addressable solely by licensed spectrum networks.

What's driving demand?

Streaming and cloud-based services and applications are the biggest demand drivers. They are enabled by better devices and richer applications and reinforced by trends to higher resolution screens and the availability of lower latency, higher performance 4G (LTE) networks. As the younger generation's unprecedented consumption of data anywhere and on any device becomes the de facto behavior in the larger populous, wireless demand will climb even faster, especially where wireline broadband is insufficient or unavailable.

This model of demand growth decouples demand (consumption desire) from supply (traffic) and allows strategic, informed decisions to be made by network operators about how much demand they can meet — and to choose optimal strategies for doing so. (See Appendix B for a more in-depth discussion of our methodology.)

Importantly, we conclude that with 3G, 4G/Long Term Evolution (LTE) and small cells alone, operators will not be able to profitably address even half of the demand left untouched by Wi-Fi-like technologies.

The current cost to deliver a megabyte of data is inconsistent with the demand and willingness/ability to pay. Balancing this equation will require faster adoption of new technologies, such as SDN, network function virtualization (NFV) and new wireless network technologies and solutions (for example, 5G), as well as new business models that support new revenue generation.

Whose desire to consume are we talking about?

While we focus on modeling human data consumption, we have also accounted for traffic generated by machine-to-machine (M2M) communications or IoT in industry segments that are expected to be the main demand drivers. Our models indicate huge growth potential for traffic generated by IoT connected devices relative to its current low base, particularly post 2020. However, for this study period (2014–2020), the human desire to consume data services is still expected to be the dominant driver of traffic. Similarly, although we anticipate that virtual reality applications will become increasingly popular for remote instruction, control, learning and exploration, and consume 100 Mb/s or more per session, we anticipate that the full impact of such services will be after 2020.

Bell Labs Consulting grouped global consumers into five age-related categories:

- Generation Z (~1–12 years)
- Millennial teens (~13–18 years)
- Millennial young adults (~19–29 years)
- Generation X (~30–54 years)
- Mature adults (~55+ years)

Each of these groups is inclined to use different applications to varying degrees based on a range of factors including device type, time of day and location (for example, home, school/work, at a hotspot, in transit), application type and technological proficiency.

Bell Labs Consulting looked at the full mix of available mobile and other wireless/Wi-Fi-enabled devices by country and operator, to evaluate which user segments prefer which devices based on application and location and categorized the devices in terms of how “smart” they are, how evolved they are, and their screen size and resolution. The mix of devices and associated technologies is as follows:

Device type

- Feature phones (basic mobile phones)
- Smartphones
- Tablets/PCs
- Larger gaming or portable TV devices

Mobile technologies

- 2G
- 3G
- 4G/LTE

While newer technologies such as 5G are emerging and some operators are planning early deployments in the 2018 timeframe, widespread adoption is not expected to occur before 2020, after the first standards are set.

Device "facts" at a glance

- People prefer smaller devices when in transit and larger devices at home.
- Teens tend to own more advanced but less expensive models of a given device.
- Small and medium-sized devices are used almost equally for audio streaming.
- Mid-sized devices are preferred about 70 percent of the time for video streaming because they offer better resolution while still being relatively portable.
- Home use device choices are driven mostly by availability, not ownership (for example, a teen may own a 3G phone but will use an adult's tablet to watch video content).
- Traffic outside the home is driven by the type of device owned and the technology available (Wi-Fi, 3G, 4G/LTE).

What are users doing with their devices?

The growing per capita ownership of devices is driven by the wider availability of large screen devices that have higher resolution cameras and greater consumption and generation of streaming media. For this study, Bell Labs Consulting identified five primary categories of mobile applications and compiled insights across all regions. The following section discusses these along with a sample of granular insights from the North America region — where usage is very high and is a reasonable predictor of usage that subsequently occurs in other regions.

Streaming

- Audio (for example, Apple Music, Spotify)
- Video (for example, Netflix, Amazon Prime, YouTube, Livestream)
- Personal video sharing (for example, Periscope, Vine, GoPro)

Key insights

- Consumption demand for streaming media will grow 17 times from 2014 levels, accounting for 71 percent of the total bandwidth demand by 2020.
- Video consumption will grow the fastest at 20 times.
- Audio consumption will increase 10 times; however, due to lower bandwidth requirements, audio will amount to just 24 percent of total demand by 2020.
- Most streaming (66–74 percent) will be home-based, driven by more content and larger and higher resolution devices, with video streaming representing the lion's share at 43–57 percent of the total streaming demand.
- Generation X users will be the heaviest consumers of streamed content, responsible for nearly half of all streaming demand, as they have the financial means to pay for more expensive data plans and higher-end devices that consume more bandwidth.
- Millennial young adults, mature adults, millennial teens and Generation Z users contribute 21 percent, 18 percent, 9 percent and 6 percent, respectively, to streaming usage at home and even more so outside of the home, although Generation X users consume 8 times more than Generation Z users and 4 times more than mature adults.
- Millennial teens are the heaviest audio users, with millennial young adults close behind, consuming 3 times more than mature adults or Generation Z.
- Bandwidth requirements for audio are independent of device size; small and medium-sized devices are almost equally preferred, whether the user is at home or away.
- For video, medium-sized devices are preferred (by up to 74 percent) because they offer better resolution, as well as acceptable portability.
- Not surprisingly, the preference for larger screens is higher at home than when users are in transit.

Computing

- Mobile device applications (for example, application downloads, application management)
- Cloud processing (for example, user interactions with the cloud-based productivity and media applications and web services)

Key insights

- Computing demand comes from cloud-based applications and native apps on mobile devices.
- In 2014, end-device apps accounted for about 84 percent of computing demand. By 2020, this percentage will shift downwards as cloud-based applications grow from 16 percent to 45 percent of the total computing demand to allow for the creation of richer and more contextual experiences.
- Generation X and millennial young adults account for 82 percent of demand, with millennial young adults being heavier users.
- Small devices generate slightly more demand in total than medium-sized devices for computing applications.
- Medium-sized devices generate roughly 90 percent of the demand for cloud processing.

Storing

- Media sharing (for example, photos, videos)
- Backup and device sync (for example, device control and management)

Key insights

- As a contributor to storage-related consumption demand, media sharing will grow from 6 percent to 48 percent — used mostly by Generation X and millennial young adults, who together account for 81 percent of storage demand, with millennial young adults having much higher use rates.
- Almost all (90 percent) of media sharing and 55 percent of backup and sync demand is generated by medium-sized devices.

Gaming

- Casual: requiring limited bandwidth and minimal user engagement
- Interactive: requiring more bandwidth, lower latency and greater user interaction
- Immersive: requiring high bandwidth, very low latency and heavy user interaction

Key insights

- The top 10 percent of users game about 3 times more than the average user.
- Mid-sized tablets are the preferred gaming devices for their balance of form factor and portability.
- Casual gaming sessions last between 10 and 30 minutes on average; immersive and interactive games tend to be played for 1.5 to 4 hours.
- Immersive gaming has the fewest users but produces higher demand due to higher bit rates.

Communicating

- Voice (for example, voice over Internet Protocol [VoIP] and voice-based conferencing)
- Video (for example, video calls, mixed video/voice conferences, large-screen video conferencing)
- Email and attachments
- Instant messaging (for example, WhatsApp, Line, Snapchat)
- IoT and wearables

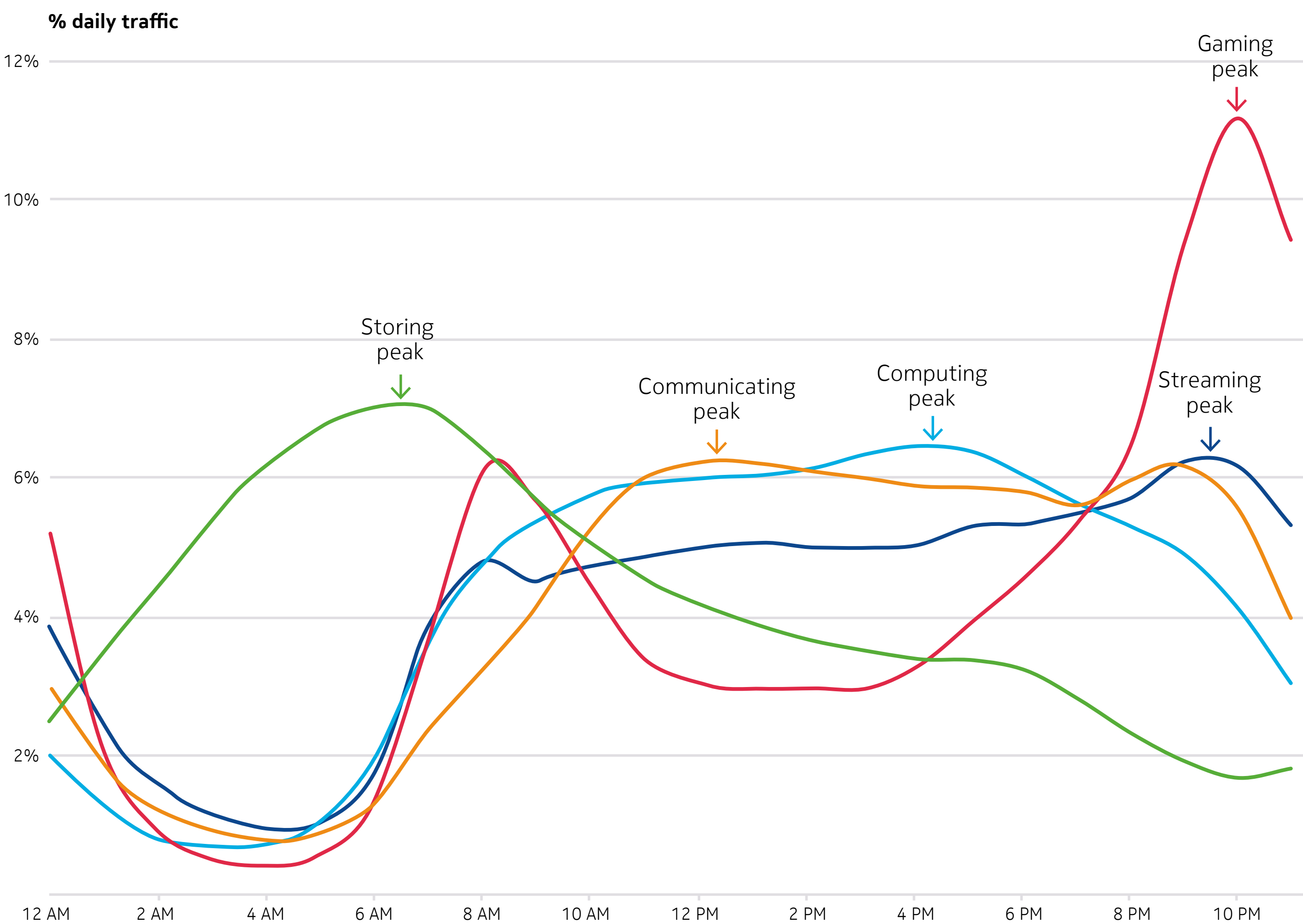
Key insights

- In this category, as in others, video contributes the most communicating demand — and will climb from 47 percent in 2014 to 86 percent in 2020, driven by millennial teens and young adults.
- As video communication rises, email demand will fall from 47 percent in 2014 to 7 percent in 2020, though the daily number of messages sent will remain constant at around 80 to 90 per user (with messaging being increasingly the driver of communications growth).
- Generation X and millennial young adults account for roughly 57 percent of all demand.
- As video and voice are the main demand drivers, the preferred devices are medium and large.

Each application category has its own usage profile, which is defined by quality of service (QoS) requirements such as latency or sustained throughput or by daily usage. The result is a different demand peak or busy hour for each application category: storage in the early morning, communicating and computing over the course of the business day, and gaming and streaming dominating the evening — with gaming peaking latest of all (see Figure 1). Modeling this usage profile helps translate business plans and marketing forecasts, usually expressed in usage per month, into network capacity requirements, expressed in Mb/s or Erlangs (simultaneous usage measure). It is these network capacity requirements that ultimately drive operators' investments.

Another crucial part of the usage profile is the location mix for each application. Consumption is very different at home, at work or in transit. Users at home have more leisure time and access to a greater number and variety of devices — leading them to consume more and better quality video content for longer periods on larger, high resolution devices. This is also important from a network planning perspective. It helps to forecast the difference in traffic growth for different network domains and to identify potential solutions that will allow delivery in the most cost-effective way. In this study, Bell Labs Consulting has modeled and forecasted usage for three location categories: at home, at work and in transit.

Figure 1. Daily traffic distribution for the five primary categories of mobile applications



Source: Bell Labs Consulting

Conservative or disruptive?

To capture the complete picture of future wireless data demand, Bell Labs Consulting has made projections along two trajectories (see Figure 2).

The first is a conservative estimate based on the premise that early adopters will primarily embrace richer and more immersive applications such as virtual reality and the majority of the user base will only gradually follow suit — accompanied by growing adoption of tablets and smartphones, increased multi-tasking and social networking and the availability of new applications. This conservative approach considers usage across all devices that have either wireless or Wi-Fi connectivity.

The second approach is more disruptive. It presumes that what looks like early adopter behavior today — for example, the use of 360° cameras or augmented/virtual reality devices — will be mainstream in the future, disrupting the steady growth assumptions of the conservative model. For the disruptive model to be realized, the ecosystem must be ready: everything from device chipsets and application availability to network performance. Operators must take fast, aggressive action to satisfy the disruptive demand.

Regional view color key

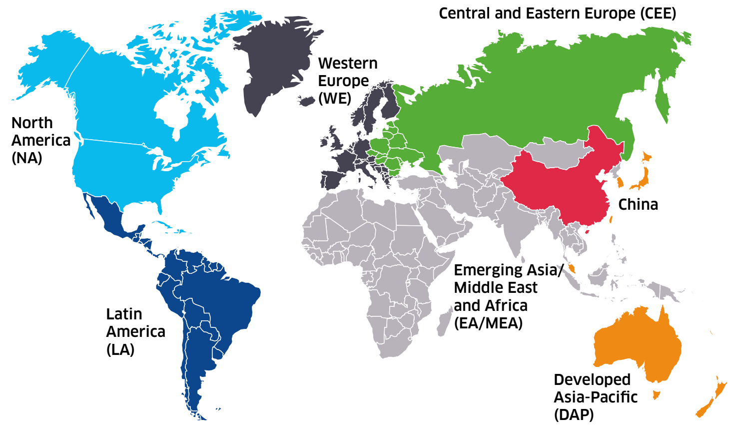
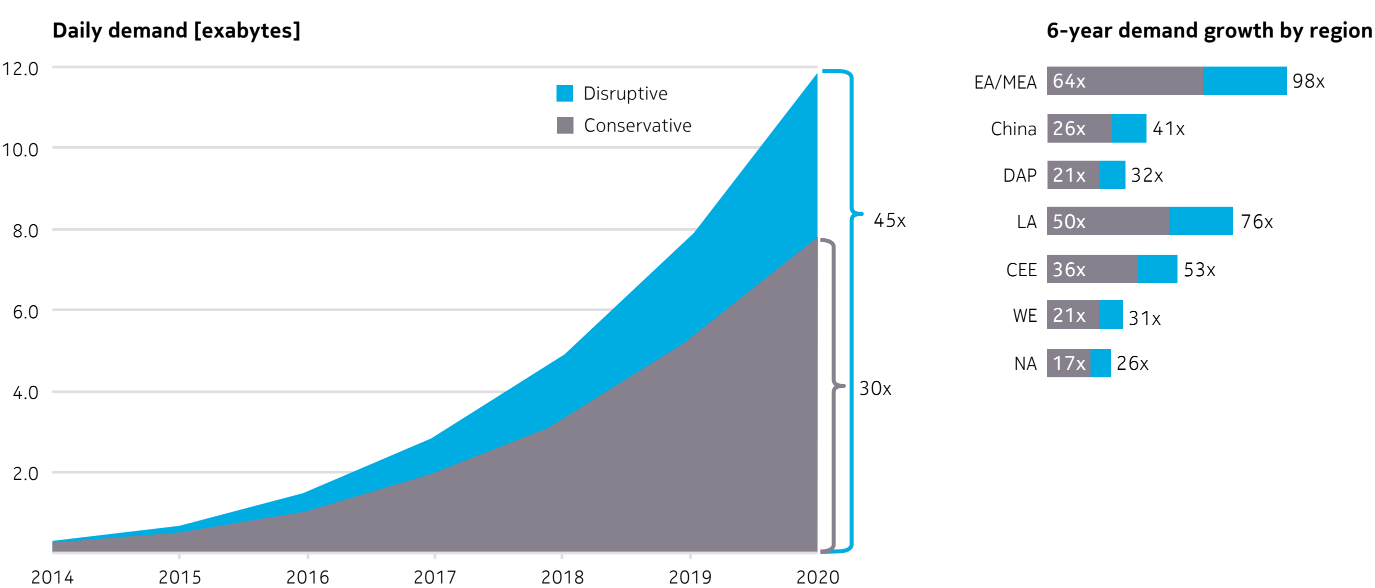


Figure 2 contrasts the two trajectories, projecting 30 times conservative growth and 45 times for disruptive growth. Although the difference between these two models is only 50 percent in 2020, the divergent trend is clear and will result in a much greater disparity beyond 2020.

Within the conservative trajectory, the 30-fold global average growth ranges from 17-fold growth in the deeply established NA market to 64-fold growth in EA/MEA markets. (See Appendix A for a breakdown of each region by country.) In the disruptive scenario, these lows and highs are both magnified — with 26-fold growth in NA and a staggering 98-fold growth in EA/MEA.

The remainder of this report focuses primarily on the conservative growth model as that is already sufficiently challenging to satisfy, as described in the following sections.

Figure 2. Global wireless data demand growth – conservative versus disruptive trajectories



Source: Bell Labs Consulting

Wireless data demand growth will be higher where wireline is lacking

Regions that lack a robust or extensive wireline infrastructure will see greater increases in wireless data demand over the coming years. In those cases, wireless will become the primary mode of broadband access, with the highest growth being driven by the uptake of mobile broadband. For operators in emerging markets, the challenge will therefore be to bring down costs and find alternative ways of meeting the massive demand growth.

Population size is a major determinant of growth in emerging markets. As seen in Figure 3, while NA has the highest share of current demand followed by China, EA/MEA will have the highest share of demand in 2020. Demand in both EA/MEA and China are expected to surpass that of NA in 2016 (see Figure 4).

Regional view color key

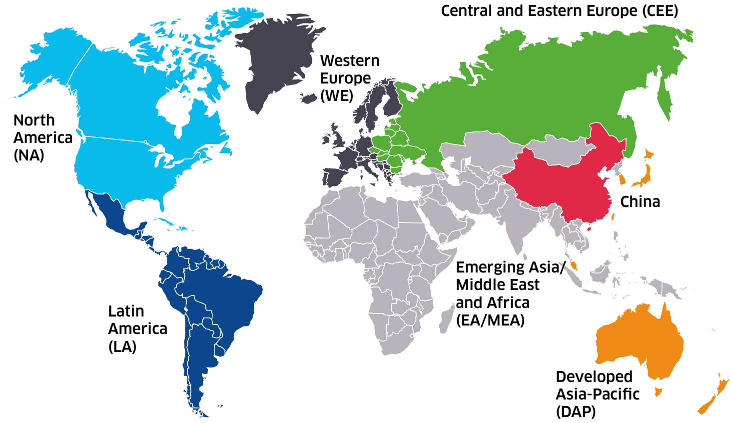
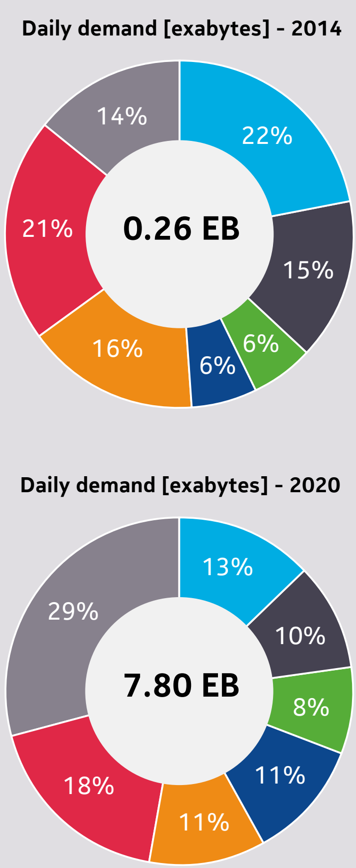
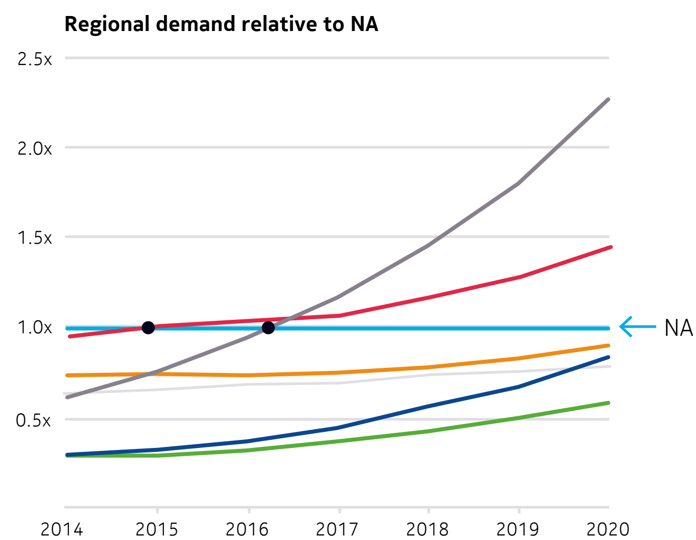


Figure 3. A regional view of demand and its growth



Source: Bell Labs Consulting

Figure 4. More people, more petabytes – consumption by population



Source: Bell Labs Consulting

Lack of wireline alternatives is another major determinant in growth of wireless data demand. As seen in Figure 5, the growth in data consumption demand per capable device is between 22 times and 33 times for those regions that do not have a robust wireline infrastructure — Central and Eastern Europe (CEE), Latin America (LA) and EA/MEA — versus 14 times and 16 times in developed regions — NA, Western Europe (WE) and Developed Asia-Pacific (DAP). However, DAP and NA will continue to see much higher data demand per device than the rest of the world.

Regional view color key

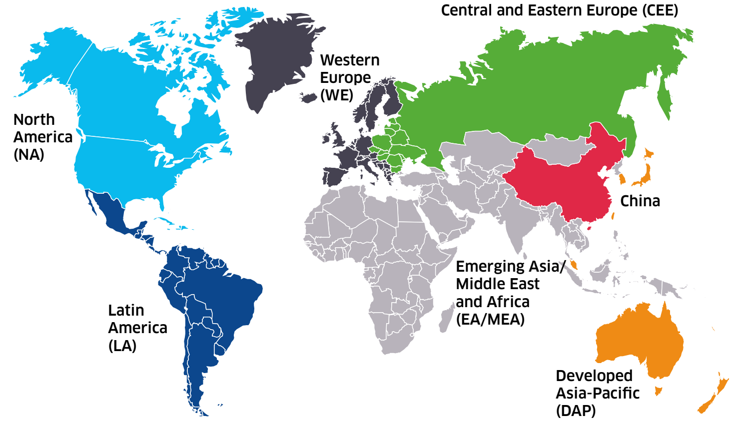
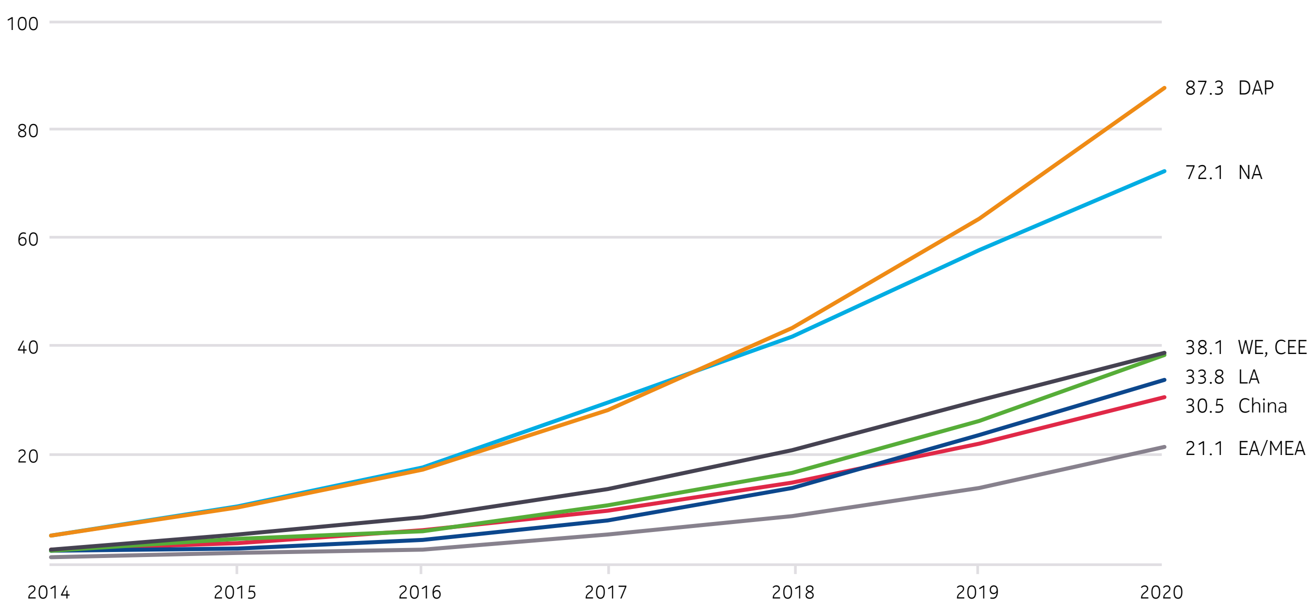


Figure 5. Growth in three dimensions

2014 to 2020 growth	NA	WE	CEE	LA	DAP	China	EA/MEA
Population	1.1x	1.0x	1.0x	1.1x	1.0x	1.0x	1.1x
Data capable devices* (smartphones, tablets)	1.2x	1.5x	1.6x	1.8x	1.3x	1.7x	2.0x
Data demand per device*	13.7x	14.3x	22.0x	27.5x	16.3x	15.0x	32.6x

* Wearables and M2M devices not included

Data demand per device [GB/month]



Source: Bell Labs Consulting

Angle on applications: streaming will dominate

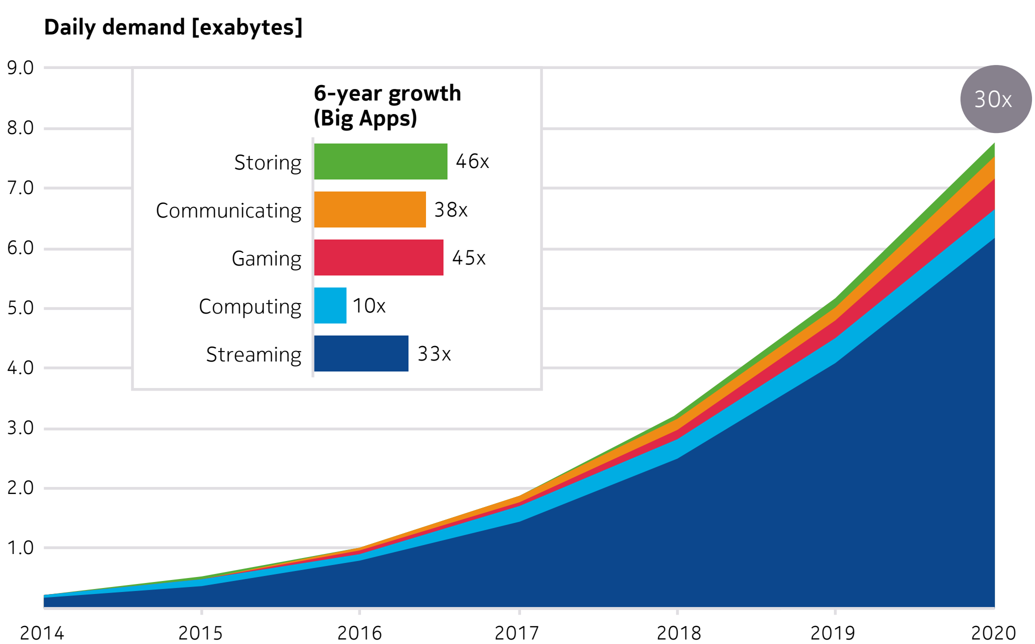
Across the five identified application categories — streaming, computing, storing, gaming and communicating — audio and video streaming will be the highest contributors to the increased traffic demand, accounting for 79 percent of demand by 2020. As mentioned previously, this will be due in part to the increased availability of better devices and richer applications driving more mobile broadband adoption and giving users larger screens on which to stream content. As discussed in the next section, IoT/machine-based video streaming will become more prevalent after 2020.

As seen in Figure 6, video applications will account for the majority of demand (56 percent), including video streaming, user-generated video and video communication. In terms of demand for apps, storage will see the most growth between now and 2020, increasing 46 times as users fully embrace the cloud and the attendant ubiquity of content and services.

Streaming dominates in 2020

Audio and video streaming = 79% of demand by 2020. Why? Better devices, content-rich applications and available networks.

Figure 6. Video dominates even though other applications grow faster



Video applications will contribute 56% of total demand

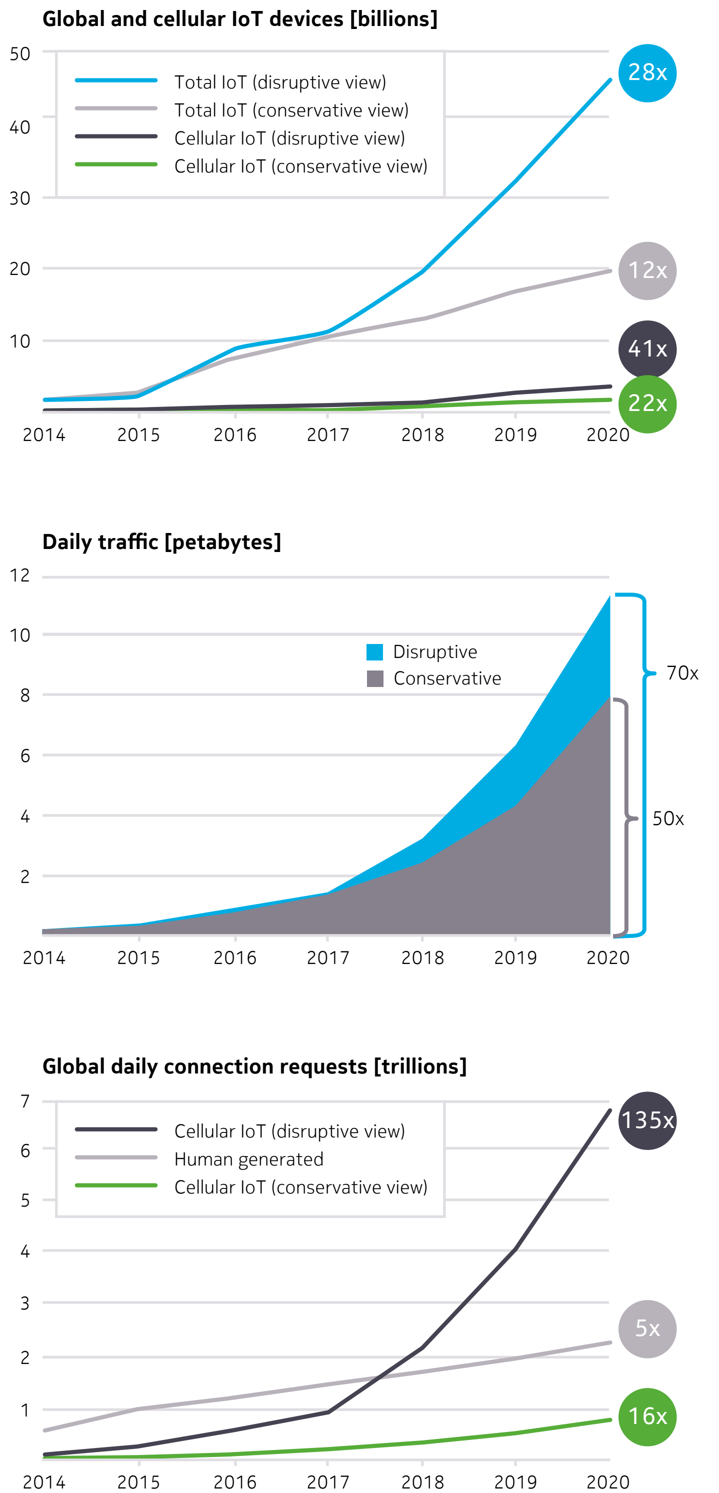
Source: Bell Labs Consulting

IoT explosion: small traffic impact but big implications on the network

Digitizing and connecting physical things to the internet (IoT) is widely predicted to occur on a massive scale in the coming decade. These connected things will use network resources to communicate with cloud-based applications and systems for the purposes of monitoring, storage, data analysis and control. The total number of IoT connected devices (not including wearables) is expected to grow from 1.6B in 2014 to anywhere between 20B (conservative view) and 46B (disruptive view) by 2020 (see Figure 7). Of this total, cellular IoT devices will be between 1.6B and 4.6B in 2020.* Despite this massive adoption and traffic growth of 50 to 70 times from 2014, the overall cellular traffic generated by IoT devices will only account for 2 percent of the total mobile traffic by 2020. The reason for this is that non-video-enabled IoT devices will predominate early on and typically transfer a small amount of data in a given transaction. However, there will be significant growth in upstream IoT video streaming after 2020 from video surveillance cameras, dashcams, body cams, and similar devices transferring content to cloud-based video analytics platforms.

With the advent of IoT, operators will also have to address the need for massive increases in control plane capacity to handle the sporadic transmissions generated from billons of devices. IoT traffic generates a substantially higher volume of signaling traffic relative to data traffic. For example, a typical IoT device may need 2,500 transactions or connections to consume 1 MB of data, while the same amount of data can be consumed in a single mobile video connection.* In the disruptive view, daily network connections due to cellular IoT devices will grow by 16 to 135-fold by 2020 and will represent 3 times the number of connections initiated by human-generated traffic.

Figure 7. IoT – small traffic impact but big implications on signaling



Source: Bell Labs Consulting

* Source: *The Future X Network: A Nokia Bell Labs Perspective*, Weldon, 2015

What it all means for mobile operators

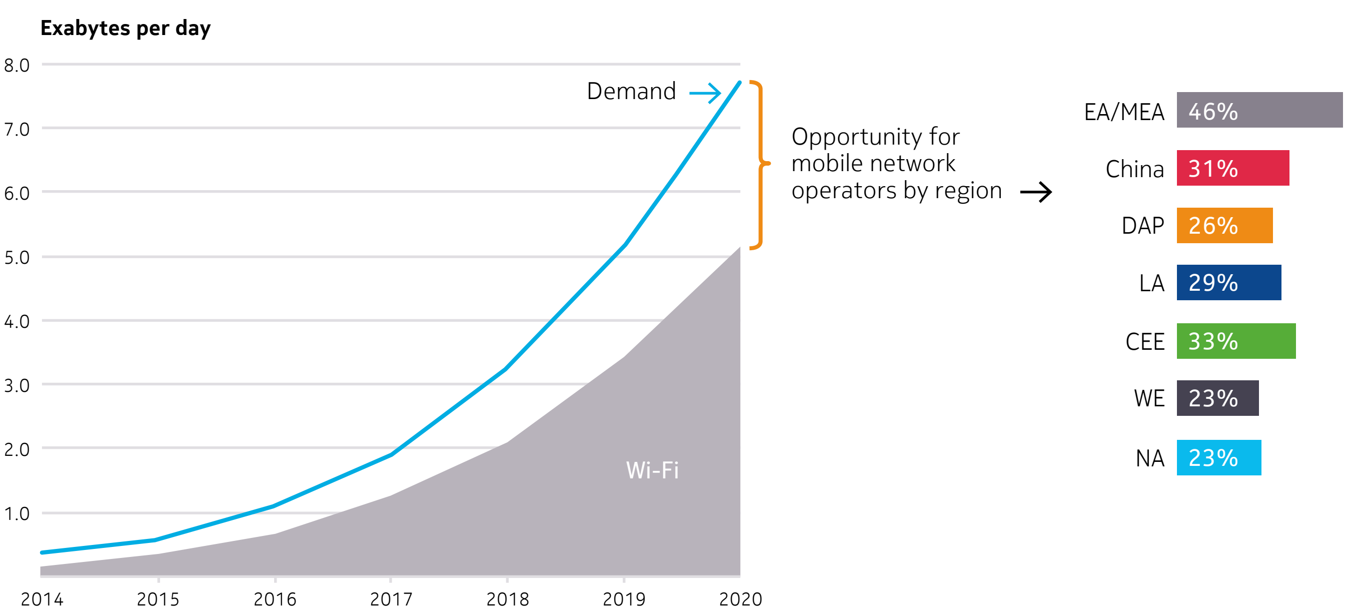
Globally, Wi-Fi already carries 60 percent of wireless traffic (Figure 8). By 2020, it will handle about 67 percent of the projected demand increase in the conservative model. Wi-Fi is most applicable in places with well-established wireline infrastructures because it is economical and fits the usage paradigm: Most Wi-Fi usage happens in fixed locations such as home, work, school and retail/public settings.

The consequence of more ubiquitous Wi-Fi is that users will become even more conditioned to expect wireless connectivity wherever they go — and not just any wireless connectivity, but broadband connectivity. These expectations present mobile operators with the opportunity to provide wireless broadband coverage wherever Wi-Fi is absent or cannot meet service requirements. This represents 33 percent of the overall demand in 2020; in emerging markets without a substantial wireline infrastructure this percentage will necessarily be substantially higher.

But this raises the key question: How much of this demand can and will be delivered economically by mobile/licensed spectrum network operators?

Users will become even more conditioned to expect wireless connectivity wherever they go — and not just any wireless connectivity, but broadband connectivity.

Figure 8. Between capacity and demand lies opportunity



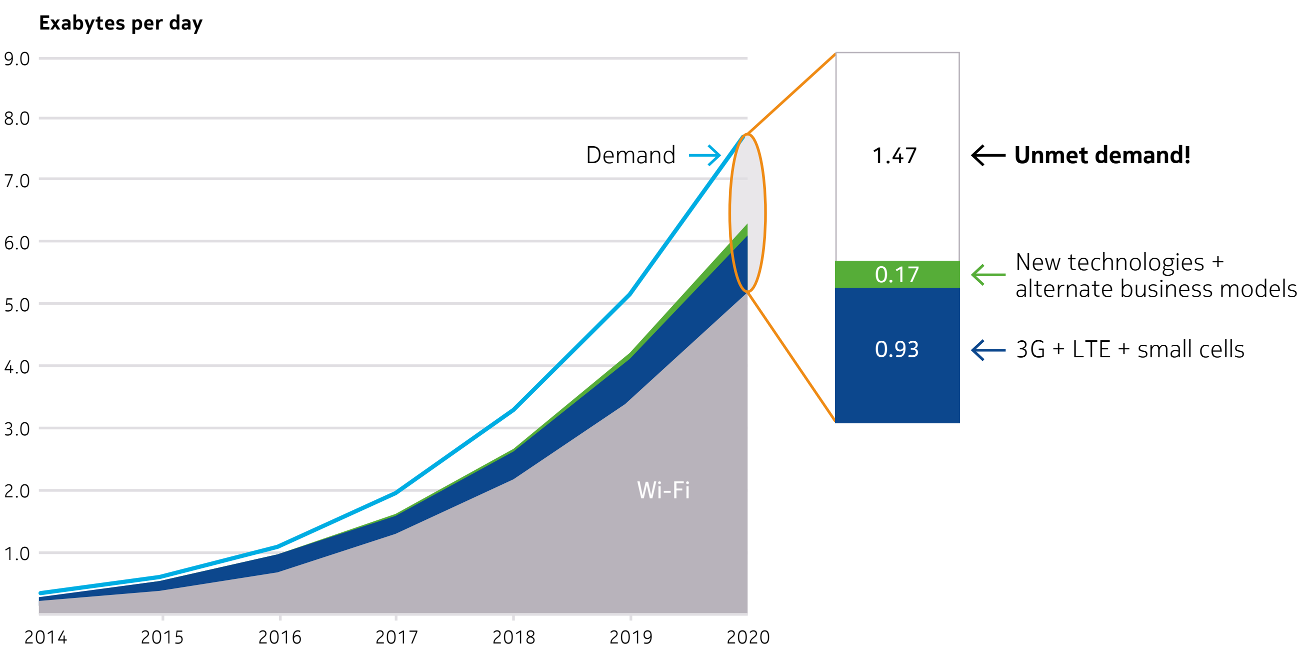
Source: Bell Labs Consulting

The boundary condition for how much demand such operators can satisfy is determined in part by the technology and architectural choices they make and the business models they employ, all of which have an impact on their operating margins. A further constraint is affordability — how much consumers are willing to spend for the value they derive from mobile data. The bottom line, in every case, is that whatever demand they address operators must do so profitably.

As shown in Figure 9, close to a third of the residual demand can be met with operator deployments of LTE and small cells along with their existing legacy 3G networks. A further 7 percent can be met by the deployment of new technologies, such as SDN, NFV and 5G, as well as the adoption of alternate business models (discussed in more detail in the next section). However, with the current trajectory of operator network evolutions, more than half of the nascent demand cannot be profitably addressed by 2020.

Region to region, the total unaddressed demand ranges from 3 percent to 36 percent; globally it averages at about 19 percent. In other words, there is a gap that needs to be addressed to fully realize the digital future.

Figure 9. Currently planned technologies and strategies can only partially address the opportunity gap



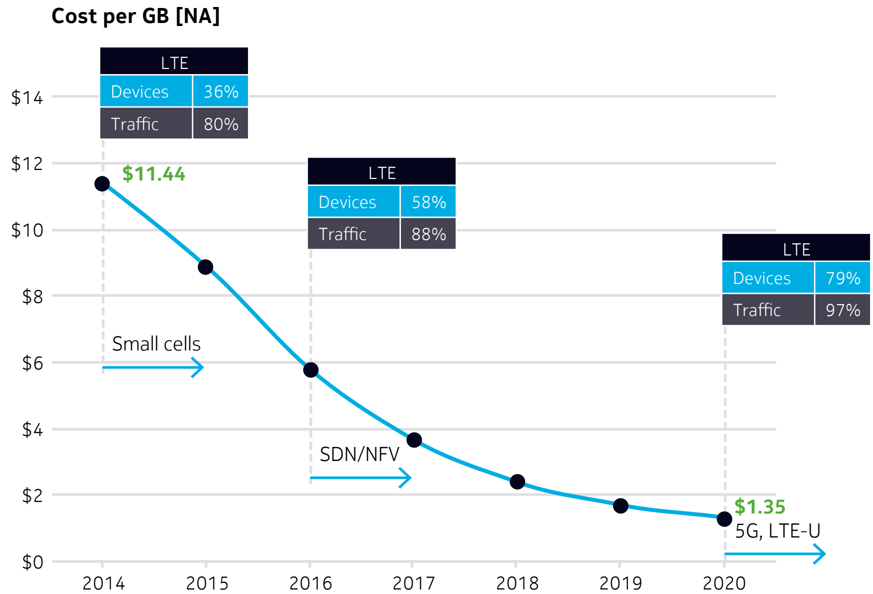
Closing the gap: the technology case

To address this unmet demand, operators will need to accelerate their path to convergence and adoption of newer technologies, such as SDN, NFV and 5G. These technologies will allow them to achieve the critical goal of delivering agile, elastic and highly personalized services, with optimized cost per bit per application at each location and time of day, enabling sustained leadership and profitability.

Every region will evolve along its own track. In NA, for example, Bell Labs Consulting has projected an expected cost per gigabit curve based on accelerated 4G/LTE adoption and scale efficiencies due to increased traffic in the network, deployment of SDN/NFV from 2016 onward, followed by further cost reductions with rollouts of 5G and Long Term Evolution unlicensed (LTE-U) after 2020. As illustrated in Figure 10, from a baseline of more than $11/GB, the cost curve traces a steady and significant decline to just $1.35/GB by 2020.

Operators are already beginning to adopt SDN/NFV and are increasingly moving toward converged networks. The degree of benefit each operator realizes will depend on how aggressively they adopt new technologies and next-generation architectures. There are clearly both conservative and disruptive paths that can be followed by operators to achieve this goal, and the path selection will likely depend on shareholder and board expectations in the near-term compared to the longer-term view.

Figure 10. New technologies rapidly push down the cost curve



Source: Bell Labs Consulting

Different technologies will yield different gains in profitable and available bandwidth for supporting mobile applications and services. Continuing with the NA example, the cumulative impact of 2G/3G/4G/LTE will deliver about 4.8 GB per user per month by 2020, supporting 4.3 times the traffic growth that was delivered in 2014. Through adoption of technological advances such as small cells, smart loading, 5G, LTE-U and SDN/NFV, operators will be able to profitably deliver about 12.8 GB per user per month, which will be 10.5 times over what was supported in 2014.

Furthermore, alternative business models, such as third-party pay models (discussed more extensively in the next section), can result in new revenue streams allowing operators to support even higher usage — up to 14.5 GB per user per month in 2020 or 12 times more than what they supported in 2014. Table 1 shows the cumulative effects of adopting several technologies and strategies.

Table 1. Mobile demand and traffic growth by strategy

Strategy	2020 average GB/month [cumulative]	Growth relative to 2014
2G/3G	0.5	1.4x
4G	4.8	4.3x
Small cells	8.6	6.8x
New technologies*	12.8	10.5x
Alternate models	14.5	12.0x

* Including smart loading, SDN/NFV, 5G and LTE-U.

Source: Bell Labs Consulting

As noted, each region will evolve along its own track for technology adoption. Thus, the ability to profitably support the bandwidth demand will also evolve along different trajectories. Figure 11 shows how the mobile data usage per device is expected to evolve during the period 2014–2020 for each region.

Regional view color key

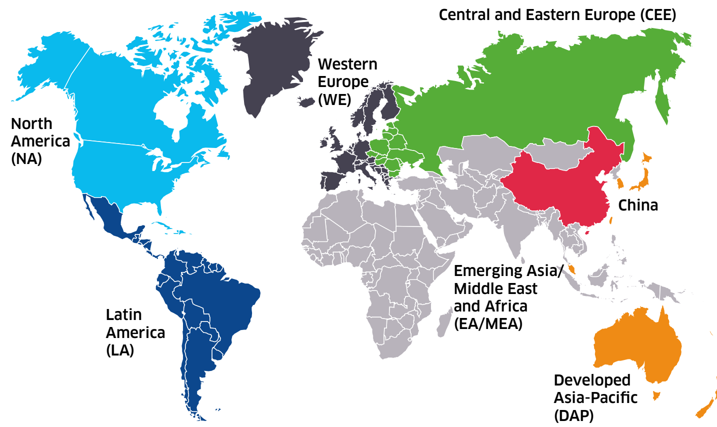
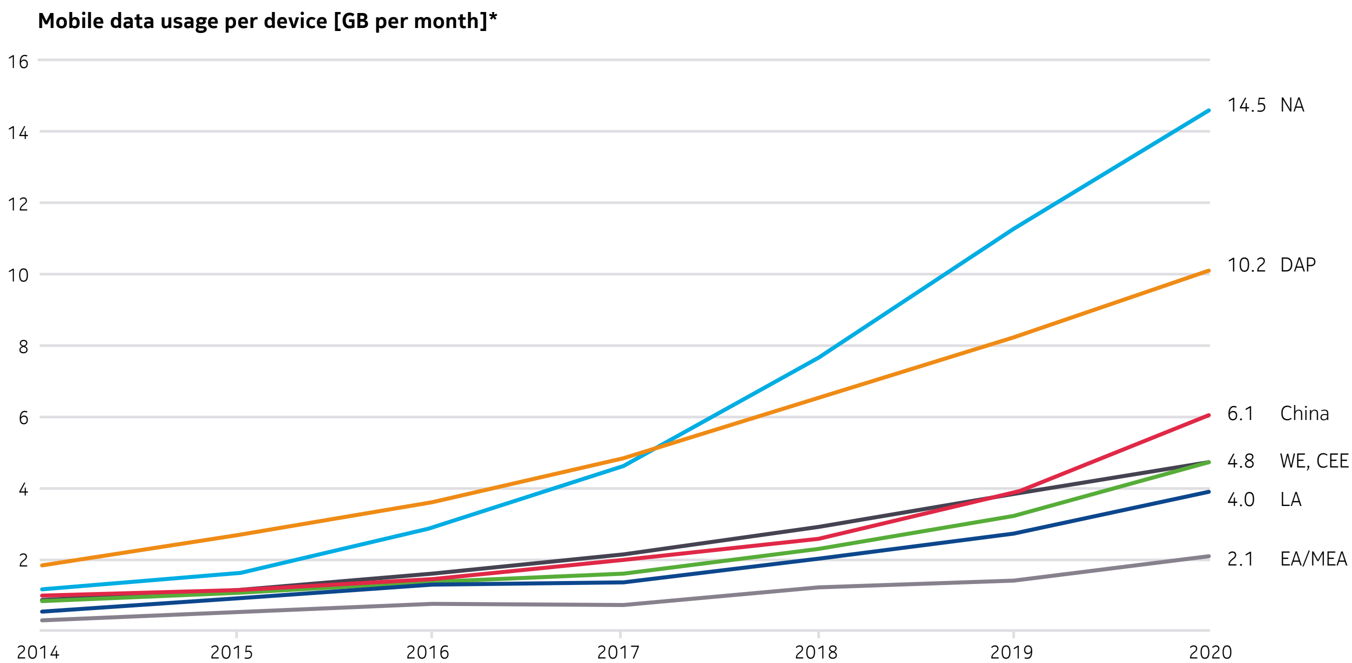


Figure 11. Profitable cellular consumption per data capable device [GB/month]



* Consumer usage (not IoT)

Source: Bell Labs Consulting

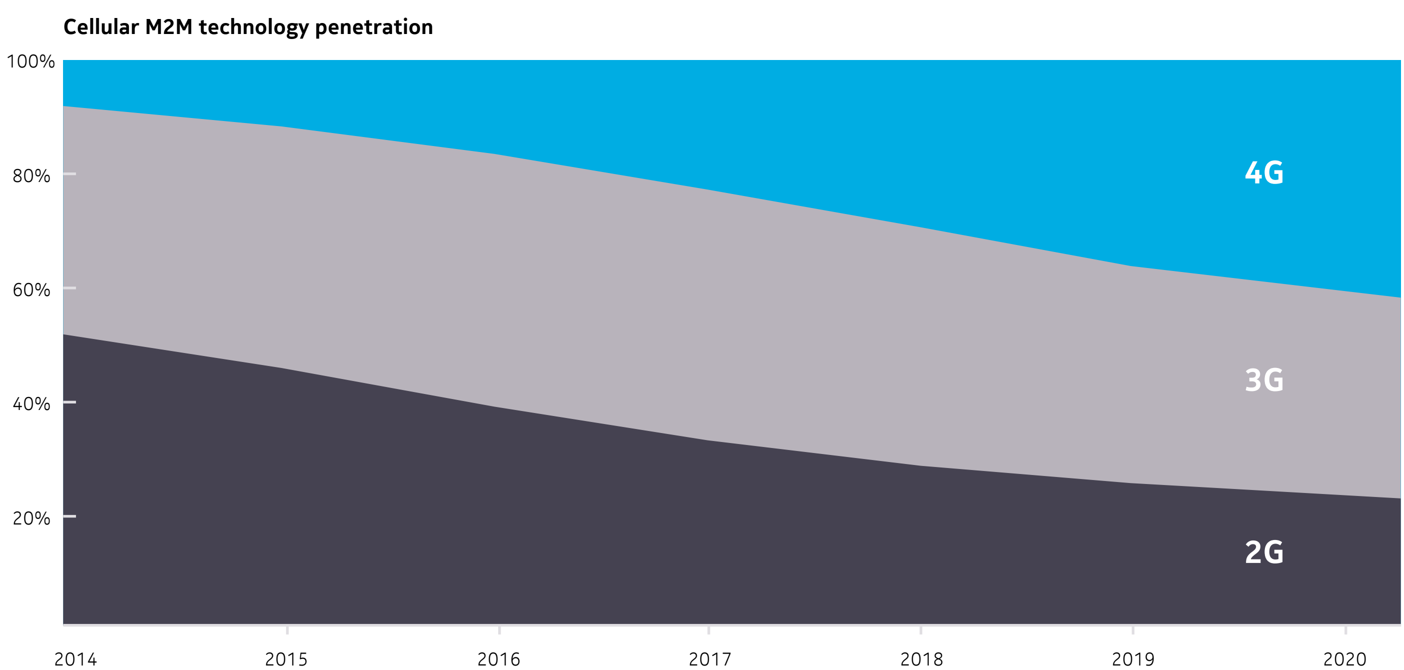
In the preceding analysis we have focused on user-generated traffic. However, as mentioned previously IoT brings a host of other challenges. With the sheer ubiquity and volume of IoT devices, contention for finite control plane network resources will be extreme, impacting other human-critical services. As a contrast to human communications, collaboration and content delivery applications, different IoT applications have different traffic profiles and also do not necessarily follow human diurnal cycles. IoT applications also have highly diverse requirements in terms of network connectivity, reliability, security, latency, data rate, mobility and battery life. These requirements must also be met at extremely lost cost per bit, due to the lower value per bit (for example, to communicate an on/off state for a "thing") than for a typical (human) cellular connection.

To address the cellular (wide area network) IoT requirements, multiple generations of wireless technologies will likely be utilized to support the lower cost, extended coverage and improved battery life for end devices. Although legacy cellular technologies, such as 2G, are not highly data efficient, they are ubiquitous and can economically support the lower data rates of simple devices. Even though operators are likely to rapidly move to 4G/LTE and the attendant IoT related enhancements such as LTE for Machine Type Communications/Narrow Band IoT (LTE-M/NB IoT), a significant portion of the IoT connectivity is still likely to be through legacy 2G — and evolutions such as Enhanced Coverage GSM (EC-GSM) — and 3G, as shown in Figure 12. Long swap cycles for IoT devices, the low cost of 2G and 3G devices, and the fact that legacy networks are fully amortized will ensure reasonable longevity of these legacy cellular technologies.

To support the growing needs of IoT, the future wireless network will have the following characteristics:

- Designed to prevent congestion and minimize impact on high average revenue per user (ARPU) human consumable services
- Cost-effective connectivity through virtualized network functions
- Enhanced air interface specifications for lower power operation and long battery life
- Flexibility and scalability to handle exponential growth and diverse IoT device and application requirements
- Dedicated network or virtualized network slices for end-to-end reliability and security
- Edge cloud architecture for ultra-low latency required by some IoT applications

Figure 12. High adoption of 4G/LTE M2M devices but 2G will stay



Source: Bell Labs Consulting

Switching up the business model

As alluded to in the preceding section, operators are already looking at alternative business models to generate revenue to compensate for increased demand needs.

One such model is sponsored data, which brings content providers to the data-pricing table. Simply put, the model allows companies to reverse-charge data consumption by buying the data from the mobile operator on behalf of end users — similar to toll-free voice numbers — in exchange for which the sponsor gains the opportunity to improve customer satisfaction, reduce transactional friction or promote their product or brand.

The actual application of the strategy can be quite diverse, from encouraging uptake of a new app or promoting a new film to delivering wellness videos to healthcare patients. In the business-to-business (B2B) market, it could include supporting bring your own device (BYOD) policies by paying for employee data usage related to specific apps and services in the consumer realm.

Facebook has a long history of sponsoring data routed through its mobile site in developing markets (Facebook "Free Basics" service). AT&T also began offering sponsored data in January 2014. Some of its partners include FreedomPop, a wireless internet and mobile phone service provider in Los Angeles, which has introduced free 100 percent sponsored data plans. Similarly, T-Mobile offers a "Binge On" service in NA that zero-rates music streaming from several music services, such as Pandora, SoundCloud, Rhapsody and many more. Syntonic Wireless is another startup platform service provider bundling content with mobile data connectivity for innovative commercial content models. DataMi offers marketplaces for content providers to sponsor data for different apps.

Another approach is to create a different business model; for example, the operator collects the data fee as part of another related transaction. The AT&T offer of unlimited music downloads with Beats Music is another example. AT&T customers get unlimited streaming music and downloads for offline listening — a deal they can extend to up to five family members across 10 devices. Each family member has 24/7 access to more than 20 million songs at a cost of $14.99 per month.

Although these models are both interesting and innovative, they are currently raising regulatory concerns either in terms of net neutrality policy or in terms of creating an underclass of connected populations in emerging economics. For example, the recent Prohibition of Discriminatory Tariffs for Data Services Regulations 2016 ruling by the Telecom Regulatory Authority of India (TRAI) stated that no company can charge subscribers with discriminatory tariffs for data services on the basis of content, which essentially eliminates the Facebook Free Basics offer in India.

The potential of such business models remains unclear at the current time, but we anticipate they will be an increasingly important part of supporting a rich digital future for everyone.

Summary and the architecture of the future

In summary, at the current rate of adoption, 3G, 4G/LTE, small cells and Wi-Fi-like technologies together with the emergence of new technologies will only satisfy about 81 percent of the conservative demand forecast by 2020. Our disruptive models indicate, however, that the demand could be 50 percent higher than the conservative forecast. If the disruptive forecast holds, operators would only be able to address 75 percent of the total demand. Operators that take steps early and aggressively to adopt new technologies and margin-sustaining business models will be able to address more of the demand gap profitably (see Figure 9).

To seize the opportunity, there are some architectural considerations mobile operators must make. More interactive and time-critical applications will become prevalent, requiring high throughput and low latency and at a lower cost per bit. It is sobering and revealing to consider that the speed of light is the key limitation for some applications; to achieve sub 1 ms latencies required for human visual perception and the so-called "haptic" control applications, the signal can only travel 100 km (less if some queuing or processing delay is included), which essentially mandates that the cloud has to move to the edge of the network, creating a distributed edge cloud with compute capacity close to the user. This will also increase throughput per user, as traffic will not traverse the more heavy, statistically multiplexed core network.

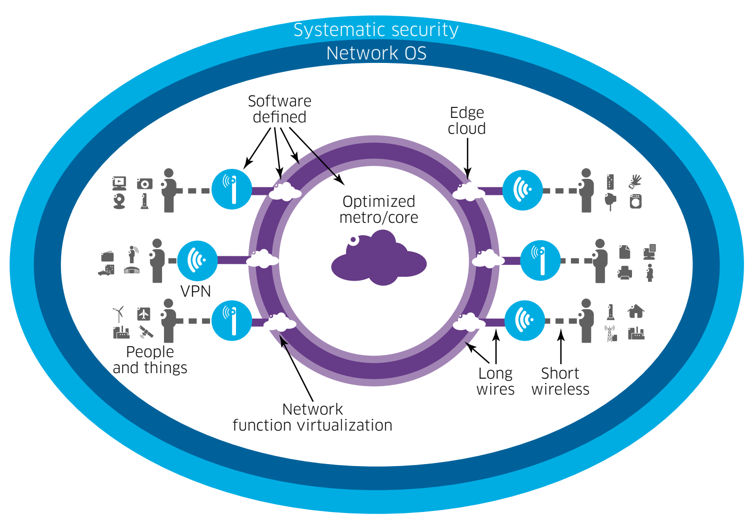
In addition, this edge cloud is required to support the virtualization of the network control plane, for which there are also critical timing requirements (~4 ms for a split virtual RAN architecture and 300 microseconds for a full cloud RAN solution). Such architectural evolutions are required to support the massively scalable control plane required by the anticipated IoT growth underlying and creating the digital future.

So, in essence both the network and the new applications it will enable mandate this shift to an edge cloud future architecture, as shown in Figure 13. This evolution is discussed in detail in the recent book, *The Future X Network: A Nokia Bell Labs Perspective.*

In conclusion, our analysis shows that the potential unmet demand can be significantly reduced or eliminated by a dramatic move to next-generation technologies and architectures. However, the cost of this transition remains a key question and one that needs to be analyzed on a case by case basis, as it is highly dependent on the current state of network infrastructure, the labor costs, rights of way and local regulatory constraints, as well as the ability for operational support systems to support legacy and future network systems and to rapidly on-board new services. This is the subject of ongoing work by Bell Labs Consulting with operators around the globe.

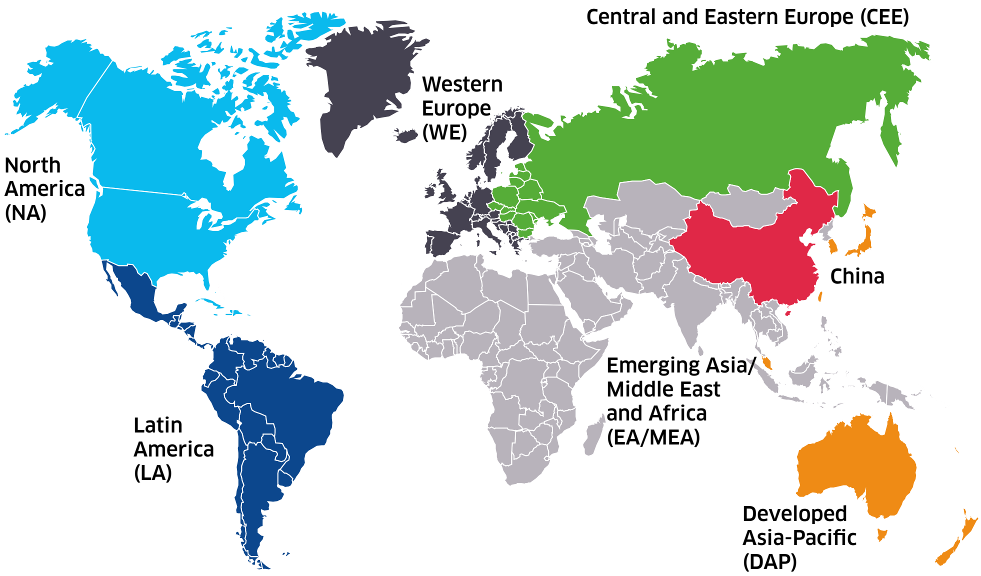
Moreover, this new mobile network architecture will pave the way for life automation and "smart everything" — an intelligent and interconnected world where technology seamlessly enables and enriches our daily lives, and where users' desires are (almost) boundlessly fulfilled, whether at home or in transit. In that future, the network will not only be something that facilitates our everyday activities, it will be inseparable from them — ushering in a brand new era where networked intelligence from everything and everywhere augments the progress of civilization.

Figure 13. Schematic depiction of the cloud-integrated network architecture



Source: *The Future X Network: A Nokia Bell Labs Perspective*

Appendix A: Regions with country listings



The following is a list of all countries considered within the regions referenced throughout this report.

North America (NA): Canada, United States

Latin America (LA): Argentina, Bahamas, Barbados, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Trinidad and Tobago, Uruguay, Venezuela

Western Europe (WE): Andorra, Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Liechtenstein, Luxembourg, Malta, Monaco, Netherlands, Norway, Portugal, San Marino, Spain, Sweden, Switzerland, United Kingdom

Central and Eastern Europe (CEE): Albania, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Estonia, Hungary, Kosovo, Latvia, Lithuania, Macedonia, Moldova, Montenegro, Poland, Romania, Russia, Serbia, Slovakia, Slovenia, Ukraine

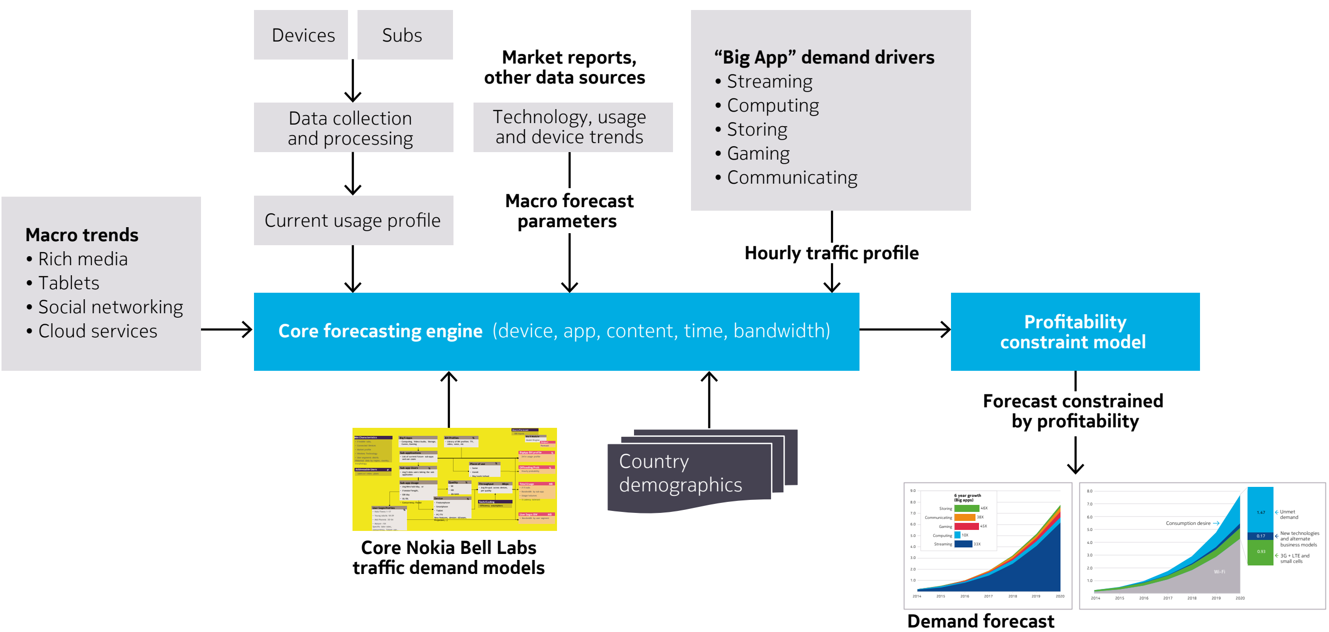
Developed Asia-Pacific (DAP): Australia, Hong Kong, Japan, New Zealand, Singapore, South Korea, Taiwan

China

Emerging Asia/Middle East and Africa (EA/MEA): Afghanistan, Algeria, Angola, Armenia, Azerbaijan, Bahrain, Bangladesh, Benin, Bhutan, Botswana, Burkina Faso, Burundi, Cambodia, Cameroon, Cape Verde, Central African Republic, Chad, China – Macau, Comoros, Democratic Republic of the Congo, Djibouti, East Timor, Egypt, Equatorial Guinea, Eritrea, Ethiopia, Federated States of Micronesia, Fiji, France – French Polynesia, France – Réunion, Gabon, Gambia, Georgia, Ghana, Guinea, Guinea – Bissau, India, Indonesia, Iran, Iraq, Israel, Ivory Coast, Jordan, Kazakhstan, Kenya, Kiribati, Kuwait, Kyrgyzstan, Laos, Lebanon, Lesotho, Liberia, Libya, Madagascar, Malawi, Malaysia, Maldives, Mali, Marshall Islands, Mauritania, Mauritius, Mongolia, Morocco, Mozambique, Myanmar, Namibia, Nauru, Nepal, New Zealand – Cook Islands, New Zealand – Niue, Niger, Nigeria, Norfolk Island, North Korea, Oman, Pakistan, Palestine, Papua New Guinea, Philippines, Qatar, Republic of the Congo, Rwanda, Sao Tome and Principe, Saudi Arabia, Senegal, Sierra Leone, Somalia, South Africa, Sri Lanka, Sudan, Swaziland, Syria, Tajikistan, Tanzania, Thailand, Togo, Tonga, Tunisia, Turkey, Turkmenistan, Tuvalu, Uganda, United Arab Emirates, United States – American Samoa, United States – Guam, Uzbekistan, Vanuatu, Vietnam, Yemen, Zambia, Zimbabwe

Appendix B: Methodology

Nokia Bell Labs mobile forecast methodology is based on a two-step approach. Initially, we estimate the end-user's desire or need to consume wireless data and then apply operators' supply side constraints to come up with a forecast of mobile data traffic. This decoupling of the demand drivers from operators' supply options allows for two things: first, the examination of how end-user demand for data will evolve as a function of individual needs and preferences, device evolution, applications and macro-economic factors; and second, how supply side constraints and network evolution strategies impact operators' ability to profitably support the given demand. The following diagram provides a brief overview of the methodology and data source linkages.



Source: Bell Labs Consulting

To develop this forecast, Nokia Bell Labs used a variety of internal and external data sources including, but not limited to:

- UN country census data
- Pyramid Research
- GSMA Intelligence
- Radicati Group, Inc.
- Ovum, MRG
- Infonetics
- comScore
- Nielsen
- eMarketer
- Strategy Analytics
- Machina Research

Overlaid on these sources are Nokia Bell Labs proprietary models: time of day behavior, application evolution, technology evolution, user behavior, and machine-to-machine device/application characterization.

For information about Bell Labs Consulting, please visit:
www.bell-labs.nokia.com/consulting